

# A Topological Approach to Meta-heuristics: Analytical Results on the BFS vs. DFS Algorithm Selection Problem

Tom Everitt and Marcus Hutter

Australian National University, Canberra, Australia

September 10, 2015

## Abstract

Search is a central problem in artificial intelligence, and BFS and DFS the two most fundamental ways to search. In this report we derive results for average BFS and DFS runtime: For tree search, we employ a probabilistic model of goal distribution; for graph search, the analysis depends on an additional statistic of path redundancy and average branching factor. As an application, we use the results on two concrete grammar problems. The runtime estimates can be used to select the faster out of BFS and DFS for a given problem, and may form the basis for further analysis of more advanced search methods. Finally, we verify our results experimentally; the analytical approximations come surprisingly close to empirical reality.

## Keywords

BFS, DFS, Analytical Algorithm Selection, Average runtime, Meta-heuristics, Tree Search, Graph Search, Probabilistic Goal Distribution

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Graph search problems</b>	<b>3</b>
2.1	Algorithm performance . . . . .	5
<b>3</b>	<b>Basic Search Algorithms</b>	<b>6</b>
3.1	Trees . . . . .	6
3.2	Uninformed search methods . . . . .	6
3.3	Informed Constructive Methods . . . . .	7
3.4	Informed Local Search . . . . .	8
<b>4</b>	<b>Literature review</b>	<b>8</b>
4.1	Feature-based Meta-heuristics . . . . .	8
4.2	Learning the search policy . . . . .	9

<b>5</b>	<b>Complete Binary Tree</b>	<b>10</b>
5.1	Preliminaries . . . . .	11
5.2	Complete Binary Tree with a Single Goal Level . . . . .	12
5.3	Complete Binary Tree with Multiple Goal Levels . . . . .	15
5.3.1	DFS Analysis . . . . .	15
5.3.2	BFS Analysis . . . . .	16
<b>6</b>	<b>Colliding Branches</b>	<b>17</b>
6.1	DFS Analysis . . . . .	18
6.2	BFS Analysis . . . . .	19
<b>7</b>	<b>Grammar Problems</b>	<b>20</b>
7.1	Binary Grammar . . . . .	20
7.2	Random Grammar . . . . .	22
7.3	Full Grammar . . . . .	23
<b>8</b>	<b>Experimental verification</b>	<b>24</b>
<b>9</b>	<b>Empirical Predictions</b>	<b>27</b>
<b>10</b>	<b>Discussion</b>	<b>29</b>
<b>A</b>	<b>List of notation</b>	<b>33</b>
<b>B</b>	<b>List of search problems</b>	<b>34</b>
<b>C</b>	<b>List of Topological features</b>	<b>35</b>

## 1 Introduction

A wide range of problems in artificial intelligence can be naturally formulated as *search problems* (Russell and Norvig, 2010; Edelkamp and Schrödl, 2012). Examples include planning, scheduling, and combinatorial optimisation (TSP, graph colouring, etc.), as well as various toy problems such as Sudoku and the Towers of Hanoi. Search problems can be solved by exploring the space of possible solutions in a more or less systematic or clever order. *Meta-heuristics* are general search methods not aimed at a specific type of problem. They interact with the problem through abstract properties such as a neighbourhood relation on the space of feasible solutions, and a heuristic or objective function. One possible way to create flexible meta-heuristics is to combine a *portfolio* of search algorithms, and use problem features to predict which search algorithms works the best. Predicting the best algorithm is sometimes known as the *algorithm selection problem* (Rice, 1975).

A number of studies have approached the algorithm selection problem with machine learning techniques (Kotthoff, 2014; Hutter et al., 2014). While demonstrably a feasible path, machine learning tend to be used as a *black box*, offering little insight into *why* a certain method works better on a given problem. On the other hand, most existing analytical results focus on worst-case big-O analysis, which is often less useful than average-case analysis when selecting algorithm. An important worst-case result is Knuth’s (1975) simple but useful technique

for estimating the depth-first search tree size. Kilby et al. (2006) used it for algorithm selection in the SAT problem. See also the extensions by Purdom (1978), Chen (1992), and Lelis et al. (2013). Analytical IDA\* runtime predictions based on problem features was obtained by Korf et al. (2001) and Zahavi et al. (2010). In this study we focus on *theoretical analysis* of *average runtime* of BFS and DFS. While the IDA\* results can be interpreted to give rough estimates for average BFS search time, no similar results are available for DFS.

To facilitate the analysis, we use a probabilistic model of goal distribution and graph structure. Currently no method to automatically estimate the model parameters is available (this is an important line of future research). Regardless, the analysis still offers important theoretical insights into BFS and DFS search. The parameters of the model can also be interpreted as a *Bayesian prior belief* about goal distribution. A precise understanding of BFS and DFS performance is likely to have both practical and theoretical value: Practical, as BFS and DFS are both widely employed; theoretical, as BFS and DFS are two most fundamental ways to search, so their properties may be useful in analytical approaches to more advanced search algorithms as well. In particular, our results may be a first step to selecting search strategy based on the (local) topology of the problem graph, an aspect ignored by most meta-heuristics.

Our main contribution is an analysis of expected BFS and DFS runtime as a function of tree depth, goal level, branching factor and path redundancy (Sections 5–7). We also verify the results experimentally (Section 8). Most of these results will be published as (Everitt and Hutter, 2015a,b). Definitions of different types of search problems and search algorithms are given in Section 2 and 3, and a review of related work can be found in Section 4. Conclusions and outlooks come in Section 9 and 10. Finally, Appendix A provides a list of notation, and Appendix B and C contain lists of search problems and potentially useful topological features.

## 2 Graph search problems

A common feature of many search problems is that there are a set of operations for cheaply modifying a proposed solution into similar proposed solutions. This makes it natural to view the problem as a *graph search problem*, where proposed solutions are *states* or *nodes*, and the modification operations induce *directed edges*.

We define two kinds of graph search problems.

**Definition 1** (Constructive graph search problem). A *constructive graph search problem* consists of a *state space*  $S$ , a starting state  $s_0 \in S$ , and the following efficiently computable functions:

1. Neighbourhood  $N : S \rightarrow 2^S$
2. Goal check  $C : S \rightarrow \{0, 1\}$
3. Edge cost:  $EC : (S \times S) \rightarrow \mathbb{R}^+$
4. Heuristic  $h : S \rightarrow \mathbb{R}^+$
5. Contract (tree/graph, solution depth, number of goals, admissible/consistent heuristic)

Items 4 and 5 are optional. A *constructive solution* is a path  $s_0, \dots, s_n$  from the starting state  $s_0$  to a goal state  $s_n$  with  $C(s_n) = 1$ . The *solution quality* of the path  $s_0, \dots, s_n$  is  $\sum_{i=0}^{n-1} EC(s_i, s_{i+1})$ .

For instance, planning problems are formalised as constructive graph search problems. The neighbourhood function gives a list of states reachable by a single action from the given state. The goal check indicates whether a state is a goal, and the edge cost indicates how costly it is to use a certain action (how it affects the solution quality). A solution is a sequence of actions leading to a goal state.

A heuristic may give an estimate of how close the given state is to a goal state (in terms of edge cost). An important class of heuristics are the *admissible* ones, that never overestimate the distance. *Consistent heuristics* additionally respect the triangle inequality. The properties of the heuristics may be given in the contract. The contract is not formalised here, but may be given as a dictionary of known problem properties.

**Definition 2** (Local graph search problem). A *local graph search problem* consists of a *state space*  $S$  together with the following efficiently computable functions:

1. Neighbourhood  $N : S \rightarrow 2^S$
2. Constraint  $C : S \rightarrow \{0, 1\}$
3. Objective function  $Q : S \rightarrow \mathbb{R}$
4. Contract (continuity properties of the objective function)

Item 4 is optional. A *local solution* is a state  $s \in S$ ,  $C(s) = 1$ , and its *solution quality* is  $Q(s)$ .

In local graph search problems, the goal is to find an  $s \in S$  that satisfies the constraints  $C$  and achieves as high objective value as possible. An objective function  $Q$  is (*Lipschitz*) *continuous with respect to the neighbourhood topology* with Lipschitz constant  $d$  if  $|Q(v_1) - Q(v_2)| < d$  whenever  $v_1$  and  $v_2$  are neighbours. Lipschitz continuity can make a problem easier, as it allows the objective values of surrounding nodes to be estimated from the current target value.

The search for an optimal circuit layout is one example of a problem that naturally formalises as a local graph search problems. Neighbours are reached by modifying the current layout (changing one connection), and the objective function incorporates the component cost and the energy efficiency of the layout. The constraint disqualifies circuits that fail the specifications. Also, any constructive search problem  $G_1 = \langle S_1, N_1, C_1, EC \rangle$  may be formulated as local search problem  $G_2 = \langle S_2, N_2, C_2, Q \rangle$ , by letting

- $S_2$  be the set of paths in  $G_1$ ,
- the objective function  $Q$  be the negative sum of the path cost,
- the constraint  $C_2$  check whether the last node of the path is a goal node, and
- the neighbourhood function  $N_2$  extend or contract a path by adding or removing a final node according to  $N_1$  (better choices of  $N_2$  may be available).

For example, the travelling salesman problem can be viewed as a *constructive* problem where a path is built step-by-step, or as a *local* problem where a full path is modified by swapping edges, and the objective function equals the summed edge cost. Some potentially useful structure may be lost in the conversion from a constructive to a local problem.

Although mixtures of local and constructive search problems are possible (e.g., combining an objective function with a constructive solution and edge cost), most practical graph search problems naturally formalises as either a constructive or a local graph search problem.

**Constraint Satisfaction Problems** Search problem can also often be naturally cast as constraint satisfaction problems (CSPs). Blum and Roli (2003) suggest that this offers a unified view of many search problems. A CSP formulation also seems largely in line with what Pearl (1988) has in mind, although Pearl is less specific.

Constructive graph problems may be formulated as a CSP by encoding the neighbourhood relation with constraints, without significant loss of structure. (If the state space is infinite, an infinite number of CSP variables may be necessary.) For other problems, like the Quadratic Assignment Problem or Eternity II<sup>1</sup>, the CSP formulation is more natural and retains more structure. More structure makes the space of policies richer, and permits more clever strategies. It can also make analytic results harder, however.

## 2.1 Algorithm performance

A *search algorithm* is an algorithm that returns a solution (a state or a path) to a graph search problem, given oracle access to the functions  $N$  and  $C$ , and possibly either  $EC$  and  $h$ , or  $Q$  (depending on the type of the search problem).

Performance on a single problem may be defined in terms of:

1. Solution quality.
2. The number of *explored states*; a state  $s$  is considered *explored* if either  $N(s)$  or  $C(s)$  has been called.
3. The running time of the algorithm.
4. The memory consumption of the algorithm (typically measured by the maximum number of states kept in memory).

To measure performance on a finite class of graph problems, average or worst-case performance may be used. For infinite classes, items 1 and 2 are typically subject to worst-case analysis (is the procedure complete/optimal), and 3 and 4 to asymptotic worst-case or average-case analysis.

We will focus on constructive search, and measure performance by the average number of explored states until in a goal is found in. In many cases the number of explored states is proportional to the actual runtime of the algorithm (state expansion is often the dominant operation during search).

Search algorithms that always finds a goal (when there is one) are called *complete*, and algorithms that always finds an optimal solution (when there is one) are called *optimal*.

---

<sup>1</sup>The Quadratic Assignment Problem is a classic combinatorial problem, and Eternity II is a famous puzzle competition by TOMY UK Ltd.

### 3 Basic Search Algorithms

A plethora of search methods have been studied for both the constructive and the local search problems. We here review only a subset of the more important ones, and refer to the books (Russell and Norvig, 2010; Edelkamp and Schrödl, 2012) for more details. First some preliminaries on trees, a fundamental structure in search analysis.

#### 3.1 Trees

A *rooted tree* is a (directed) graph with a *root*  $s_0$  where every pair of nodes is connected by exactly one path. The *level* of a node  $v$  is the distance from the root  $s_0$  to  $v$ . The *depth*  $d$  is the length of a longest path starting from  $s_0$ . If every node on level less than  $D \in \mathbb{N}$  has exactly  $b$  children, and nodes on level  $D$  are *leaves* (have no children), then the tree is *complete with branching factor  $b$  and depth  $D$* . Such a tree will have  $b^D$  leaves and  $(b^{D+1} - 1)/(b - 1)$  nodes. In particular, *complete binary trees* (with branching factor 2) have  $2^D$  leaves and  $2^{D+1} - 1$  nodes.

A node  $v$  is a *descendant* of a node  $u$  if there is a path from  $u$  to  $v$  (i.e., if  $v$  is a child of a child of ... of  $u$ ).

#### 3.2 Uninformed search methods

*Uninformed search* refers to the case where neither a heuristic function nor an objective function is used for guidance of the search. The two standard methods for exploring a graph in this case are *Breadth-first Search (BFS)* and *Depth-first Search (DFS)*. BFS searches a successively growing neighbourhood around the the start node, while DFS follows a single path as long as possible, and *backtracks* when stuck. Algorithm 1 and 2 give pseudo-code for BFS and DFS respectively, and Figure 1 shows the traversal order of BFS and DFS in a complete binary tree.

---

**Algorithm 1** Pseudo-code for BFS

---

```
Q ← emptyQueue
Discovered ← emptySet
Q.add(start-node)
Discovered.add(start-node)
while Q not empty do
     $u \leftarrow Q.pop()$ 
    if  $C(u)$  then return  $u$ 
    for  $v$  in  $N(u)$  do
        if not  $v \in \text{Discovered}$  then
            Q.add( $v$ )
            Discovered.add( $v$ )
```

---

DFS is substantially more memory-efficient than BFS:  $O(d)$  compared to  $O(b^d)$ . However, BFS can be emulated by an *iterative deepening DFS*, with the same memory cost as DFS and only a small penalty in runtime in most graphs (in graphs with exponentially growing neighbourhoods, to be precise).

---

**Algorithm 2** Pseudo-code for (Recursive) DFS

---

```
function DFS-REC( $N, C, u$ , Discovered)
  Discovered.add( $u$ )
  if  $C(u)$  then return  $u$ 
  for  $v$  in  $N(u)$  do
    if not  $v \in$  Discovered then
      DFS( $N, C, v$ , Discovered)
Discovered  $\leftarrow$  emptySet
DFS-REC( $N, C$ , start-node)
```

---

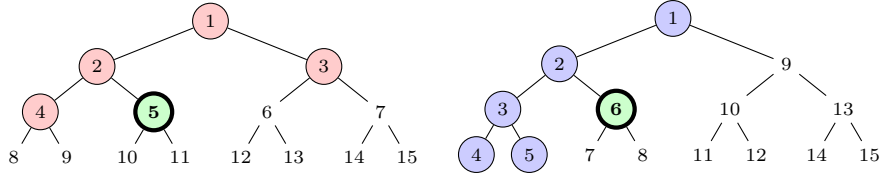


Figure 1: The difference between BFS (left) and DFS (right) in a complete binary tree where a goal is placed in the second position on level 2 (the third row). The numbers indicate traversal order. Circled nodes are explored before the goal is found. Note how BFS and DFS explore different parts of the tree. In bigger trees, this may lead to substantial differences in search performance.

BFS and DFS come in two flavors, depending on whether they keep track of visited nodes or not. The *tree search* variants do not keep track of visited nodes, while the *graph search* variants do. In trees (where each node can only be reached from one path), nothing is gained by keeping track of visited nodes. In contrast, keeping track of visited nodes can benefit search performance greatly in multiply connected graphs. Keeping track of visited nodes may also be expensive in terms of memory consumption, however.

One way to understand tree search behaviour in general graphs is to say that tree search algorithms *effectively* explore a tree; branches in this tree correspond to paths in the original graph, and copies of the same node  $v$  will appear in several places of the tree whenever  $v$  can be reached through several paths. DFS tree search may search forever if there are cycles in the graph. We always assume that path lengths are bounded by a constant  $D$ .

Depending on the positions of the goals in the graph, DFS and BFS may have substantially different performance. In Sections 5–7 below, We investigate some simple models of how the goal position and the graph structure affect BFS and DFS performance.

### 3.3 Informed Constructive Methods

*Informed constructive search methods* make use of a heuristic function. This often speeds up the search significantly. The most popular method for informed constructive search is  $A^*$ . It may be seen as a generalisation of BFS.  $A^*$  combines the heuristic information  $h(v)$  of a node  $v$  with the *accumulated edge cost*  $g(v)$  of the shortest found path from the start node to  $v$ .  $A^*$  always expands the discovered node with the smallest  $g(v) + h(v)$ . But when the heuristic function is

smaller for nodes closer to the goal,  $A^*$  will prioritise more promising nodes and find the goal much faster than BFS. If in particular the heuristic is consistent (never overestimates distance and respects the triangle inequality), then  $A^*$  is guaranteed to find an optimal solution. The tree search version of  $A^*$  only requires the heuristic to be admissible (never overestimate search distance) for guaranteed optimality.

One of the main draw-backs of  $A^*$  is its memory consumption. *Iterative deepening  $A^*$  ( $IDA^*$ )* uses a cutoff value  $c$  that is incremented between iterations. In each iteration, nodes  $v$  not satisfying  $g(v) + h(v) < c$  are ignored, while the rest are expanded in DFS manner. The memory consumption thus becomes on par with DFS. However, unlike iterative deepening BFS, the runtime slowdown of  $IDA^*$  compared to  $A^*$  may be exponentially worse, since it may only be possible to increase the cutoff  $c$  marginally between iterations if optimality of the solution must be guaranteed. A range of other memory-efficient versions of  $A^*$  can be found in the literature.

If the edge cost is always 1 and the heuristic entirely uninformative  $h \equiv 0$ , then  $A^*$  reduces to BFS and  $IDA^*$  to iterative-deepening DFS.

In many cases the priority is to find a goal as fast as possible, and the goal does not need to be an optimal one. In these cases, a *greedy best-first* search strategy may be used, that expands nodes according to  $h(v)$  instead of  $g(v) + h(v)$ . Greedy best-first may be seen as the most natural generalisation of DFS to the informed constructive scenario. *Beam-search* is a variant of greedy best-first that searches slightly more widely.

### 3.4 Informed Local Search

Most informed local search algorithms strive to combine an exploiting, hill-climbing component with an exploration component. The simplest one is *hill-climbing*, which always goes to the neighbour with the highest objective value, and randomly restarts when stuck. More advanced methods include *simulated annealing*, which adds a random moves to hill-climbing. The randomness component decays over time. *Genetic algorithms* use a *population* of search nodes, and tries to find new search points by combining features of discovered ones.

## 4 Literature review

We divide our review of related work into two subsections. The works in the first subsection assumes that a portfolio of predefined algorithms is given, and only tries to predict which algorithm in the portfolio is better for which problem. The second subsection reviews approaches that try to build new search policies, possibly using a set of basic algorithms as building blocks.

### 4.1 Feature-based Meta-heuristics

The *algorithm selection problem* asks what algorithm best to use on a given problem (Rice, 1975; Kotthoff, 2014). Tightly related is the question of inferring the search time of different search algorithms on the problem, as this information can be used to select the fastest algorithm. Both analytical investigations and machine learning techniques applied to empirical data have been tried. The latter



is sometimes known as *empirical performance models*. The most comprehensive surveys are given by Hutter et al. (2014) and Kotthoff (2014), and the PhD theses Thompson (2011) and Arbelaez Rodriguez (2011).

For DFS, Knuth (1975) made a simple but important observation how the branching factor seen during search can be used to estimate the size of the search tree and the runtime. Despite the simplicity of the scheme, the estimates work surprisingly well in practice. Several generalisations have been developed (Purdom, 1978; Chen, 1992). Kilby et al. (2006) generalise Knuth’s method, and also use it to select search policy for the SAT problem based on which search policy has the lowest estimated runtime. Kullmann (2008) and Lelis et al. (2013) both develop estimation schemes for branch-and-bound algorithms. Haim and Walsh (2008) approach the SAT problem, and instead of branching factor use properties of the given formula (such as the number and the size of clauses) to be predict search time and best search policy.

In the case of informed search, Korf et al. (2001) developed an interesting analytic technique for estimating the search time of IDA\*. Assuming a consistent heuristic function, the estimate is based on the distribution of heuristic function values at different depths of the search tree, rather than heuristic accuracy. Intuitively, the scheme works because the number of nodes expanded in each iteration of IDA\* depends on the number of nodes with heuristic value less than the threshold. The distribution of heuristic values is also easy to estimate in practice. Zahavi et al. (2010) generalise the work of Korf et al. to non-consistent heuristics.

Many other approaches instead try to directly infer the best search policy, without the intermediate step of estimating runtime. Fink (1998) does this for STRIPS-like learning using only the problem size to infer which method is likely to be more efficient. Schemes using much wider ranges of problem properties are applied to CSPs in (Thompson, 2011; Arbelaez Rodriguez, 2011), and to the NP-complete problems SAT, TSP and Mixed integer programming in (Hutter et al., 2014). Smith-Miles and Lopes (2012) review and discuss commonly used features for the algorithm selection problem, mainly applied to the local search scenario. They divide features into two main categories: General and problem-specific. General features usually phrased in terms of the *fitness landscape* (i.e., the target function and the neighbourhood structure). A common fitness landscape feature is for example the variability (*ruggedness*) of the target function. Another general feature is the performance of a simple, fast algorithm such as gradient descent. Problem-specific features are discussed for a range of NP-complete problems such as TSP and Bin-packing.

## 4.2 Learning the search policy

Explanation-based Learning (EBL) (Dejong and Mooney, 1986; Mitchell et al., 1986; Minton, 1988) is a general method for learning from examples and domain knowledge. In the context of search, the domain knowledge is the neighbourhood function (or the consequence of applying an ‘action’ to a state). An example to learn from can be the search trace of an optimiser. The EBL learner analyses the different decisions represented in the search trace, judges whether they were good or bad, and tries to find the *reason* they were good or bad. Once a reason has been found, the gained understanding can be used to pick similar good decisions at an earlier point during the next search, and to avoid similar bad

decisions (decisions leading to paths where no goal will be found). EBL systems have been applied to STRIPS-like planning scenarios (Minton, 1988, 1990).

One characteristic feature of EBL is that it requires only one or a few training examples (in addition to the domain knowledge). While attractive, it can also lead to overspecific learning (Minton, 1988). Partial Evaluation (PE) is an alternative learning method that is more robust in this respect, with less dependency on examples (Etzioni, 1993). Leckie and Zukerman (1998) develops a more inductive way to learn search control knowledge (in contrast to the deductive generalisations performed by EBL and PE), where plenty of training examples substitute for domain knowledge.

A more modern approach is known as *hyper heuristics* (Burke et al., 2003, 2013). It views the problem of inferring good search policies more abstractly. Rather than interacting with the neighbourhood structure/graph problem directly, the hyper heuristic only has access to a set of search policies for the original graph problem. The search policies are known as *low-level heuristics* in this literature (not to be confused with *heuristic functions*). The goal of the hyper heuristic is to find a good policy for when to apply which low-level heuristic. For example, Ross et al. (2002) used Genetic Algorithms to learn which bin-packing heuristic to apply in which type of state in a bin-packing problem. The learned hyper heuristic outperformed all the provided low-level heuristics used by themselves. In applications of hyper heuristics, the low-level heuristics are typically simple search policies provided by the human programmers, although nothing prevents them from being arbitrarily advanced meta-heuristics. Some research is also being done on automatic construction of low-level heuristics (see (Burke et al., 2013) for references). A related approach directed at programming in general is *programming by optimisation* (Hoos, 2012), where machine learning techniques are used to find the best algorithm in a space of programs delineated by the human programmer.

## 5 Complete Binary Tree

In a search graph, the neighbourhood relation  $N$  induces a topology on the state space  $S$ . The following two sections analytically explore how the structure and depth of the graph and the distribution of the goals can be used to predict the search performance of BFS and DFS. Figure 1 gives the intuition for the different search strategies BFS and DFS, and how they initially focus the search on different areas of the tree.

As a concrete example, consider the search problem of solving a Rubik’s cube. There is an upper bound  $D = 20$  to how many moves it can take to reach the goal (Rokicki and Kociemba, 2013). We may however suspect that most goals are located around level 17 ( $\pm 2$  levels). If we consider search algorithms that do not remember where they have been, the search space becomes a complete tree with fixed branching factor 9. What would be the expected BFS and DFS search time for this problem? Which one would be faster?

After some initial background in the first subsection, this section first investigates a model where goals are located on a single goal level  $g$  in Section 5.2, and then generalises it to multiple goal levels in Section 5.3. Section 6 develops techniques for analyzing the performance of the graph search variants of BFS and DFS, which recognize the path redundancies often present in problems. All

analytical runtime estimates are verified experimentally in Section 8.

## 5.1 Preliminaries

For simplicity, we say that the *runtime* or *search time* of a search method (BFS or DFS) is the number of nodes explored until a first goal is found (5 and 6 respectively in Figure 1). This simplifying assumption relies on node expansion being the dominant operation, consuming similar time throughout the tree. If no goal exists, the search method will explore all nodes before halting. In this case, we define the runtime as the number of nodes in the search problem plus 1 (i.e.,  $2^{D+1}$  in the case of a binary tree of depth  $D$ ).<sup>2</sup>

Let  $\Gamma$  be the event that a goal exists,  $\Gamma_k$  the event that a goal exists on level  $k$ , and  $\bar{\Gamma}$  and  $\bar{\Gamma}_k$  their complements. Let  $F_k = \Gamma_k \cap (\bigcap_{i=0}^{k-1} \bar{\Gamma}_i)$  be the event that level  $k$  has the *first* goal.

A random variable  $X$  is *geometrically distributed*  $\text{Geo}(p)$  if  $P(X = k) = (1-p)^{k-1}p$  for  $k \in \{1, 2, \dots\}$ . The interpretation of  $X$  is the number of trials until the first success when each trial succeeds with probability  $p$ . Its cumulative distribution function (CDF) is  $P(X \leq k) = 1 - (1-p)^k$ , and its *average* or *expected value*  $\mathbb{E}[X] = 1/p$ . A random variable  $Y$  is *truncated geometrically distributed*  $X \sim \text{TruncGeo}(p, m)$  if  $Y = (X \mid X \leq m)$  for  $X \sim \text{Geo}(p)$ , which gives

$$P(Y = k) = \begin{cases} \frac{(1-p)^{k-1}p}{1-(1-p)^m} & \text{for } k \in \{1, \dots, m\} \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbb{E}[Y] = \mathbb{E}[X \mid X \leq m] = \frac{1 - (1-p)^m(pm + 1)}{p(1 - (1-p)^m)}.$$

Let  $\text{tc}(p, m)$  denote the expected value of a truncated geometrically distributed variable with parameters  $p$  and  $m$  (i.e.,  $\text{tc}(p, m) = \mathbb{E}[Y]$ ). When  $p \gg \frac{1}{m}$ ,  $Y$  is approximately  $\text{Geo}(p)$ , and  $\text{tc}(p, m) \approx \frac{1}{p}$ . When  $p \ll \frac{1}{m}$ ,  $Y$  becomes approximately uniform on  $\{1, \dots, m\}$  and  $\text{tc}(p, m) \approx \frac{m}{2}$ .

A random variable  $Z$  is *exponentially distributed*  $\text{Exp}(\lambda)$  if  $P(Z \leq z) = 1 - e^{-\lambda z}$  for  $z \geq 0$ . The expected value of  $Z$  is  $\frac{1}{\lambda}$ , and the probability density function of  $Z$  is  $\lambda e^{-\lambda z}$ . An exponential distribution with parameter  $\lambda = -\ln(1-p)$  might be viewed as the continuous counterpart of a  $\text{Geo}(p)$  distribution. We will use this approximation in Section 5.3.

**Lemma 3** (Exponential approximation). *Let  $Z \sim \text{Exp}(-\ln(1-p))$  and  $X \sim \text{Geo}(p)$ . Then the CDFs for  $X$  and  $Z$  agree for integers  $k$ ,  $P(Z \leq k) = P(X \leq k)$ . The expectations of  $Z$  and  $X$  are also similar in the sense that  $0 \leq \mathbb{E}[X] - \mathbb{E}[Z] \leq 1$ .*

*Proof.* For  $z > 0$ ,  $P(Z \leq z) = 1 - \exp(z \ln(1-p)) = 1 - (1-p)^z$ , and  $P(X \leq z) = 1 - (1-p)^{\lfloor z \rfloor}$ . Thus, for integers  $k > 0$ ,  $P(Z \leq k) = P(X \leq k)$  which proves the first statement. Further,  $1 - (1-p)^{\lfloor z \rfloor} \leq 1 - (1-p)^z < 1 - (1-p)^{\lfloor z \rfloor + 1}$ , so  $P(X \leq z) \leq P(Z \leq z) < P(X \leq \lfloor z \rfloor + 1)$ . Hence  $\mathbb{E}[X] \geq \mathbb{E}[Z] > \mathbb{E}[X - 1] = \mathbb{E}[X] - 1$ , which proves the second statement.  $\square$

<sup>2</sup>It may have seem more justified to set the non-goal case to the exact number of nodes instead of adding 1. However, adding 1 makes most expressions slightly more elegant, and does not affect the results in any substantial way.

We will occasionally make use of the convention  $0 \cdot \text{undefined} = 0$ , and often expand expectations by conditioning on disjoint events:

**Lemma 4.** *Let  $X$  be a random variable and let the sample space  $\Omega = \dot{\bigcup}_{i \in I} C_i$  be partitioned by mutually disjoint events  $C_i$ . Then  $\mathbb{E}[X] = \sum_{i \in I} P(C_i) \mathbb{E}[X \mid C_i]$ .*

## 5.2 Complete Binary Tree with a Single Goal Level

Consider a binary tree of depth  $D$ , where solutions are distributed on a single goal level  $g \in \{0, \dots, D\}$ . At the goal level, any node is a goal with iid probability  $p_g \in [0, 1]$ . We will refer to this kind of problems as *(single goal level) complete binary trees with depth  $D$ , goal level  $g$  and goal probability  $p_g$*  (Section 5.3 generalises the setup to multiple goal levels).

The probability that a goal exists is  $P(\Gamma) = P(\Gamma_g) = 1 - (1 - p_g)^{2^g}$ . If a goal exists, let  $Y$  be the position of the first goal at level  $g$ . Conditioned on a goal existing,  $Y$  is a truncated geometric variable  $Y \sim \text{TruncGeo}(p_g, 2^g)$ . When  $p_g \gg 2^{-g}$  the goal position  $Y$  is approximately  $\text{Geo}(p_g)$ , which makes most expressions slightly more elegant. This is often a realistic assumption, since if  $p \not\gg 2^{-g}$ , then often no goal would exist.

**Proposition 5** (BFS runtime Single Goal Level). *Let the problem be a complete binary tree with depth  $D$ , goal level  $g$  and goal probability  $p_g$ . When a goal exists and has position  $Y$  on the goal level, the BFS search time is*

$$\begin{aligned} t_{\text{SGL}}^{\text{BFS}}(g, p_g, Y) &= 2^g - 1 + Y, \text{ with expectation} \\ t_{\text{SGL}}^{\text{BFS}}(g, p_g \mid \Gamma_g) &= 2^g - 1 + \text{tc}(p_g, 2^g) \approx 2^g - 1 + \frac{1}{p_g}. \end{aligned}$$

*In general, when a goal does not necessarily exist, the expected BFS search time is*

$$t_{\text{SGL}}^{\text{BFS}}(g, p_g) = P(\Gamma) \cdot (2^g - 1 + \text{tc}(p_g, 2^g)) + P(\bar{\Gamma}) \cdot 2^{D+1} \approx 2^g - 1 + \frac{1}{p_g}.$$

*The approximations are close when  $p_g \gg 2^{-g}$ .*

*Proof.* When a goal exists, BFS will explore all of the top of the tree until depth  $g - 1$  (that is,  $2^{(g-1)+1} = 2^g$  nodes) and  $Y$  nodes on level  $g$ , before finding the first goal. That is,  $t_{\text{SGL}}^{\text{BFS}}(D, g, p_g, Y) = 2^g - 1 + Y$ , with expected value  $2^g - 1 + \text{tc}(p_g, 2^g)$ .

In the general case, the expected value of the search time  $X$  expands as

$$\begin{aligned} \mathbb{E}[X] &= P(\Gamma) \cdot \mathbb{E}[X \mid \Gamma] + P(\bar{\Gamma}) \cdot \mathbb{E}[X \mid \bar{\Gamma}] \\ &= P(\Gamma) \cdot t_{\text{SGL}}^{\text{BFS}}(D, p, p_g \mid \Gamma_g) + P(\bar{\Gamma}) \cdot 2^{D+1} \\ &= P(\Gamma) \cdot (2^g - 1 + \text{tc}(p_g, 2^g)) + P(\bar{\Gamma}) \cdot 2^{D+1}. \end{aligned}$$

When  $p_g \gg 2^{-g}$ , then  $\Gamma \approx 1$ ,  $\bar{\Gamma} \approx 0$  and  $Y \approx \text{Geo}(p)$  which justifies the approximation.  $\square$

A memory-efficient tree-search variant of BFS can be implemented as iterative deepening DFS (ID-DFS). The runtime of ID-DFS is about twice the runtime of BFS; our results are only marginally affected by this. Korf et al. (2001) study

IDA\*, which may be seen as a generalised ID-DFS. Their Theorem 1 give a similar result to our Proposition 5 by setting: the heuristic  $h = 0$ , the number of  $i$ -level nodes  $N_i = 2^i$ , the equilibrium distribution  $P(x) = 1$ , the edge cost = 1, and the cost bound  $c$  equal to our max depth  $D$ . Their bound then comes out as  $t_{\text{SGL}}^{\text{BFS}}(g) = 2^{g+1} - 1$ , and as  $t_{\text{SGL}}^{\text{BFS}}(g) \approx 2^{g+2}$  after iteration over all levels  $\leq g$ . This corresponds to the *worst case* in our scenario.

**Proposition 6.** *Consider a complete binary tree with depth  $D$ , goal level  $g$  and goal probability  $p_g$ . When a goal exists and has position  $Y$  on the goal level, the DFS search time is approximately*

$$\begin{aligned} \tilde{t}_{\text{SGL}}^{\text{DFS}}(D, g, p_g, Y) &:= (Y - 1)2^{D-g+1} + 2, \text{ with expectation} \\ \tilde{t}_{\text{SGL}}^{\text{DFS}}(D, g, p_g \mid \Gamma_g) &:= \left(\frac{1}{p_g} - 1\right) 2^{D-g+1} + 2. \end{aligned}$$

When  $p_g \gg 2^{-g}$ , the expected DFS search time when a goal does not necessarily exist is approximately

$$\tilde{t}_{\text{SGL}}^{\text{DFS}}(D, g, p_g) := P(\Gamma)((\text{tc}(p_g, 2^g) - 1)2^{D-g+1} + 2) + P(\bar{\Gamma})2^{D+1} \approx \left(\frac{1}{p_g} - 1\right) 2^{D-g+1}.$$

*Proof.* One way to count the nodes explored by DFS when a goal exists is the following. To the left of the first goal on level  $g$ , DFS will explore  $2(Y - 1)$  subtrees rooted at level  $g + 1$ . These subtrees will have depth  $D - (g + 1)$ , and contain  $2^{D-g} - 1$  nodes each. DFS will also explore  $Y$  nodes on level  $g$  and their parents, which amounts to about  $2Y$  nodes. Summing the contributions up gives the DFS search time approximation  $\tilde{t}_{\text{SGL}}^{\text{DFS}}(D, g, p_g, Y) = 2(Y - 1) \cdot (2^{D-g} - 1) + 2Y = (Y - 1)2^{D-g+1} + 2$ .

By Lemma 4, the expected value of the search time  $X$  expands as

$$\begin{aligned} \mathbb{E}[X] &= P(\Gamma) \cdot \mathbb{E}[X \mid \Gamma] + P(\bar{\Gamma}) \cdot \mathbb{E}[X \mid \bar{\Gamma}] \\ &= P(\Gamma) \cdot \mathbb{E}[\tilde{t}_{\text{SGL}}^{\text{DFS}}(D, g, p_g, Y) \mid \Gamma] + P(\bar{\Gamma}) \cdot 2^{D+1} \\ &= P(\Gamma) \cdot ((\text{tc}(p_g, 2^g) - 1)2^{D-g+1} + 2) + P(\bar{\Gamma}) \cdot 2^{D+1} \end{aligned}$$

where the last step uses that  $(Y \mid \Gamma) \sim \text{TruncGeo}(p_g, 2^g)$ . When  $p_g \gg 2^{-g}$ , then  $\Gamma \approx 1$ ,  $\bar{\Gamma} \approx 0$  and  $Y \approx \text{Geo}(p_g)$  which justifies the approximation.  $\square$

Figure 2 shows the runtime estimates as a function of goal level.

**Comparison BFS vs. DFS.** Figure 6 on page 23 shows how the expected search time varies with goal depth (and also compares the results with the Binary Grammar Problem described in Section 7.1). The runtime estimates can be used to predict whether BFS or DFS will be faster, given the parameters  $D$ ,  $g$ , and  $p_g$ , as stated in the next Proposition.

**Proposition 7.** *Let  $\gamma_{p_g} = \log_2(\text{tc}(p_g, 2^g) - 1) / 2 \approx \log_2\left(\frac{1-p_g}{p_g}\right) / 2$ . Given the approximation of DFS runtime of Proposition 6, BFS wins in expectation in a complete binary tree with depth  $D$ , goal level  $g$  and goal probability  $p_g$  when*

$$g < \frac{D}{2} + \gamma_{p_g}$$

*and DFS wins in expectation when  $g > \frac{D}{2} + \gamma_{p_g} + \frac{1}{2}$ .*

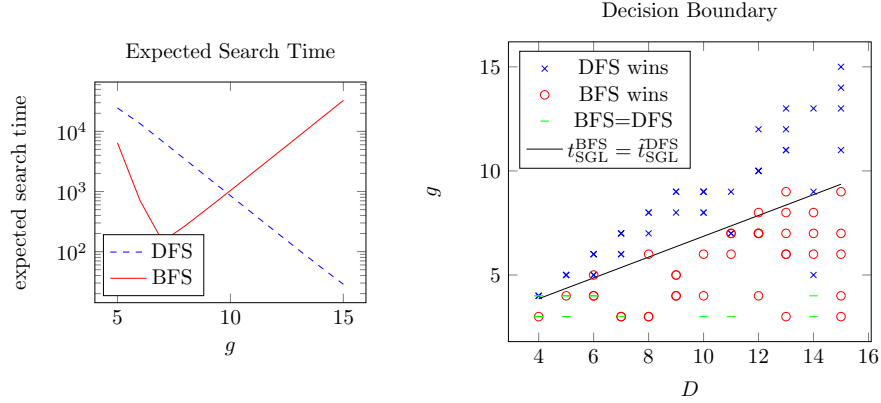


Figure 2: Two plots of how expected BFS and DFS search time varies in complete binary tree with a single goal level  $g$  and goal probability  $p_g = 0.07$ . The left depicts search time as a function of goal level in a tree of depth 15. BFS has the advantage when the goal is in the higher regions of the graph, although at first the probability that no goal exists heavily influences both BFS and DFS search time. DFS search time improves as the goal moves downwards since the goal probability is held constant. The right graph shows the decision boundary of Proposition 7, together with 100 empirical outcomes of BFS and DFS search time according to the varied parameters  $g \in [3, D] \cap \mathbb{N}$  and  $D \in [4, 15] \cap \mathbb{N}$ . The decision boundary gets 79% of the winners correct.

The term  $\gamma_{p_g}$  is in the range  $[-1, 1]$  when  $p_g \in [0.2, 0.75]$ ,  $g \geq 2$ , in which case Proposition 7 roughly says that BFS wins (in expectation) when the goal level  $g$  is located higher than the middle of the tree. For smaller  $p_g$ , BFS benefits with the boundary level being shifted  $\gamma_{p_g} \approx k/2$  levels from the middle when  $p_g \approx 2^{-k} \gg 2^{-g}$ . Figure 2 illustrates the prediction as a function of goal depth and tree depth for a fixed probability  $p_g = 0.07$ .

*Proof of Proposition 7.* When no goal exists, BFS and DFS will perform the same. When the tree contains at least one goal node, BFS will found the goal somewhere on its sweep across level  $g$ , so the BFS runtime is bounded between  $2^g \leq t_{\text{SGL}}^{\text{BFS}}(g, p_g) \leq 2^{g+1}$ .

The upper bound for  $t_{\text{SGL}}^{\text{BFS}}(g, p_g)$  gives that  $t_{\text{SGL}}^{\text{BFS}}(g, p_g) < t_{\text{DFS}}(D, g, p_g)$  when  $2^{g+1} < (tc(p_g, 2^g) - 1) 2^{D-g+1}$ . Taking the binary logarithm of both sides yields

$$g + 1 < \log_2(tc(p_g, 2^g) - 1) + D - g + 1.$$

Collecting the  $g$ 's on one side and dividing by 2 gives the desired bound

$$g < \frac{\log_2(tc(p_g, 2^g) - 1)}{2} + \frac{D}{2} = \frac{D}{2} + \gamma_{p_g}.$$

Similar calculations with the lower bound for  $t_{\text{SGL}}^{\text{BFS}}(g, p_g)$  gives the condition for  $t_{\text{SGL}}^{\text{DFS}}(D, g, p_g) < t_{\text{SGL}}^{\text{BFS}}(g, p_g)$  when  $g > \frac{D}{2} + \gamma_{p_g} + \frac{1}{2}$ .  $\square$

It is straightforward to generalise the calculations to arbitrary branching factor  $b$ , by just substituting the 2 in the base of  $t_{\text{SGL}}^{\text{BFS}}$  and  $t_{\text{SGL}}^{\text{DFS}}$  for  $b$ . In Proposition 7, the change only affects the base of the logarithm in  $\gamma_{p_g}$ .

**Corollary 8.** *Given the above approximations to BFS and DFS runtime, BFS wins in expectation in a complete tree with integer branching factor  $b \geq 2$ , depth  $D$ , goal level  $g$ , and goal probability  $p_g$  when  $g < \frac{D}{2} + \gamma_{b,p_g}$ , and DFS wins in expectation when  $g > \frac{D}{2} + \gamma_{b,p_g} + \frac{1}{2}$ , where  $\gamma_{b,p_g} = \log_b(\text{tc}(p_g, b^g) - 1)/2 \approx \log_b(\frac{1-p_g}{p_g})/2$ .*

### 5.3 Complete Binary Tree with Multiple Goal Levels

We now generalise the model developed in the previous section to problems that can have goals on any number of levels. For each level  $k \in \{0, \dots, D\}$ , let  $p_k$  be the associated *goal probability*. Not every  $p_k$  should be equal to 0. Nodes on level  $k$  have iid probability  $p_k$  of being a goal. We will refer to this kind of problems as *(multi goal level) complete binary trees with depth  $D$  and goal probabilities  $p$* .

#### 5.3.1 DFS Analysis

Our approximation of DFS performance in the case of multiple goal levels approximates the geometric distribution used in Proposition 6 with an exponential distribution (its continuous approximation by Lemma 3).

**Proposition 9** (Expected Multi Goal Level DFS Performance). *Consider a complete binary tree of depth  $D$  with goal probabilities  $\mathbf{p} = [p_0, \dots, p_D] \in [0, 1]^{D+1}$ . If for at least one  $j$ ,  $p_j \gg 2^{-j}$ , and for all  $k$ ,  $p_k \ll 1$ , then the expected number of nodes DFS will search is approximately*

$$\tilde{t}_{\text{MGL}}^{\text{DFS}}(D, \mathbf{p}) := 1 / \sum_{k=0}^D \ln(1 - p_k)^{-1} 2^{-(D-k+1)}.$$

The proof constructs for each level  $k$  an exponential random variable  $X_k$  that approximates the search time before a goal is found on level  $k$  (disregarding goals on other levels). The minimum of all  $X_k$  then becomes an approximation of the search time to find a goal on any level. The approximations use exponential variables for easy minimisation.

*Proof.* The proof uses two approximations. First approximate the position of the first goal on level  $k$  with  $Y_k \sim \text{Exp}(\lambda_k)$ , where  $\lambda_k = -\ln(1 - p_k)$ . This is reasonable for the following reason. For  $p_k \gg 2^{-k}$ ,  $Y_k$  is approximately  $\text{Geo}(p_k)$ , so the approximation is justifiable by Lemma 3. For smaller  $p_k$ , the probability that this level has a goal is small, so the imprecision of the approximation does not affect the result significantly (as long as not all  $p_j$  are this small).

Second, disregarding goals on levels other than  $k$ , the total number of nodes that DFS needs to search before reaching a goal on level  $k$  is approximately  $X_k \sim \text{Exp}(\lambda_k 2^{-(D-k+1)})$ . This follows from an approximation of Proposition 6: The number of nodes DFS needs to search to find a goal on level  $k$  is

$$\tilde{t}_{\text{SGL}}^{\text{DFS}}(D, k, p_k, Y_k) = (Y_k - 1)2^{D-k+1} + 2 \approx Y_k \cdot 2^{D-k+1}.$$

(This is a reasonable estimate if  $Y_k$  is large, which is likely given that  $p_k \ll 1$  by assumption.) So  $X_k$  is approximately a multiple  $2^{D-k+1}$  of  $Y_k$ . For any exponential random variable  $Z$  with parameter  $\lambda$ , the scaled variable  $m \cdot Z$  is  $\text{Exp}(\lambda/m)$ . This completes the justification of the second approximation.

The result now follows by a standard minimisation of exponential variables. Since  $X_k$  approximates the number of nodes searched before finding a goal on level  $k$ , the number of nodes searched before finding a goal on any level is  $X = \min_k X_k$ . The CDF for  $X$  is

$$\begin{aligned} P(X \leq y) &= 1 - \prod_{k=0}^D P(X_k > y) \\ &= 1 - \prod_{k=0}^D \exp(-\lambda_k 2^{-(D-k+1)} y) \\ &= 1 - \exp(-y \sum_{k=0}^D \lambda_k 2^{-(D-k+1)}). \end{aligned}$$

(The minimum of exponential variables  $Z_k \sim \text{Exp}(\xi_k)$  is again an exponential variable  $\text{Exp}(\sum \xi_k)$ .)

So  $X \sim \text{Exp}(\sum_{k=0}^D \lambda_k 2^{-(D-k+1)})$  with the claimed expected value  $1 / \sum_{k=0}^D \lambda_k 2^{-(D-k+1)}$ .  $\square$

In the special case of a single goal level, the approximation of Proposition 9 is similar to the one given by Proposition 6. When  $\mathbf{p}$  only has a single element  $p_j \neq 0$ , the expression  $\tilde{t}_{\text{MGL}}^{\text{DFS}}$  simplifies to

$$\tilde{t}_{\text{MGL}}^{\text{DFS}}(D, \mathbf{p}) = \frac{1}{\lambda_j} 2^{D-j+1} = -\frac{1}{\ln(1-p_j)} 2^{D-j+1}.$$

For  $p_j$  not close to 1, the factor  $-1/\ln(1-p_j)$  is approximately the same as the corresponding factor  $1/p_j - 1$  in Proposition 6 (the *Laurent expansion* is  $-1/\ln(1-p_j) = 1/p_j - 1/2 + O(p_j)$ ).

### 5.3.2 BFS Analysis

The corresponding expected search time  $t_{\text{MGL}}^{\text{BFS}}(D, \mathbf{p})$  for BFS requires less insight and can be calculated exactly by conditioning on which level the first goal is. The resulting formula is less elegant, however. The same technique cannot be used for DFS, since DFS does not exhaust levels one by one.

The probability that level  $k$  has the first goal is  $P(F_k) = P(\Gamma_k) \prod_{j=0}^{k-1} P(\bar{\Gamma}_j)$ , where  $P(\Gamma_i) = (1 - (1 - p_i)^{2^i})$ . The expected BFS search time gets a more uniform expression by the introduction of an extra *hypothetical level*  $D+1$  where all nodes are goals. That is, level  $D+1$  has goal probability  $p_{D+1} = 1$  and  $P(F_{D+1}) = P(\bar{\Gamma}) = 1 - \sum_{k=0}^D P(F_k)$ .

**Proposition 10** (Expected Multi Goal Level BFS Performance). *The expected number of nodes  $t_{\text{MGL}}^{\text{BFS}}(p)$  that BFS needs to search to find a goal in a complete binary tree of depth  $D$  with goal probabilities  $\mathbf{p} = [p_0, \dots, p_D]$ ,  $\mathbf{p} \neq \mathbf{0}$ , is*

$$t_{\text{MGL}}^{\text{BFS}}(\mathbf{p}) = \sum_{k=0}^{D+1} P(F_k) t_{\text{SGL}}^{\text{BFS}}(k, p_k \mid \Gamma_k) \approx \sum_{k=0}^{D+1} P(F_k) \left( 2^k + \frac{1}{p_k} \right)$$

For  $p_k = 0$ , the expression  $t_{\text{CB}}^{\text{BFS}}(k, p_k)$  and  $1/p_k$  will be undefined, but this only occurs when  $P(F_k)$  is also 0.



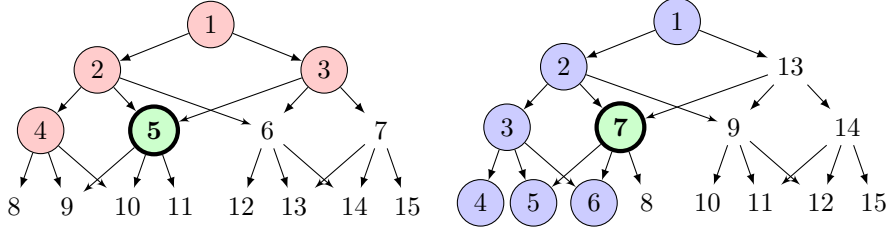


Figure 3: The difference between BFS (left) and DFS (right) in multiply connected graphs. Note how DFS is additionally concentrated at the bottom of the graph compared to Figure 1.

*Proof.* To BFS, the event  $F_k$  that level  $k$  has a goal is equivalent to the single goal level model of Section 5.2. Let  $X$  be BFS search time, and let  $(X | F_k)$  be the number of nodes that BFS needs to search when  $k$  is the first level with a goal. Then  $(X | F_k) = t_{\text{SGL}}^{\text{BFS}}(k, p_k, X - (2^k - 1) | \Gamma_k)$ , and  $\mathbb{E}[X | F_k] = t_{\text{SGL}}^{\text{BFS}}(k, p_k | \Gamma_k)$ . The result follows by expanding  $\mathbb{E}[X]$  over  $F_0, \dots, F_{D+1}$  as in Lemma 4.  $\square$

The approximation tends to be within a factor 2 of the correct expression, even when  $p_k < 2^{-k}$  for some or all  $p_k \in \mathbf{p}$ . The reason is that the corresponding  $P(F_k)$ 's are small when the geometric approximation is inaccurate.

Both Proposition 9 and 10 naturally generalise to arbitrary branching factor  $b$ . Although their combination does not yield a similarly elegant expression as Proposition 7, they can still be naively combined to predict the BFS vs. DFS winner (Figure 8).

## 6 Colliding Branches

The last section predicted runtime of *tree search* algorithms that do not remember which nodes they visit, which means that the search graph always has shape of a tree (with the same node possibly occurring in several places). In this section, we explore the performance of *graph search* algorithms that avoid revisiting previously explored nodes by keeping track of which nodes have already been seen. Figure 3 gives an idea of the difference between BFS and DFS in multiply connected graphs (with bounded search depth).

**Definition 11.** For a given search problem: Let the *level* of a node  $v$ ,  $\text{level}(v)$ , be the length of a shortest path from the start node to  $v$ . Let  $D = \max_v \text{level}(v)$  be the (*generalised*) *depth* of the search graph. Let  $\delta_n$  be the first node on level  $n$  reached by DFS,  $0 \leq n \leq D$ .

The *descendant counter*  $L$  plays a central role in the analysis. For a given search problem, let

$$L(n, d) = |\{v : \text{level}(v) = d, v \in \text{descendants}(\delta_n)\}|$$

count the number of nodes on level  $d$  that are reachable from  $\delta_n$ .

As in the previous section, we assume that goals are distributed by level in an iid manner according to a goal probability vector  $\mathbf{p}$ . We will also assume that the probability of DFS finding a goal before finding  $\delta_D$  is negligible. We will refer to

this kind of problems as *search problems with depth  $D$ , goal probabilities  $\mathbf{p}$  and descendant counter  $L$* . The rest of this section justifies the following proposition.

**Proposition 12.** *The DFS and BFS runtime of a search problem can be roughly estimated from the descendant counter  $L$ , the depth  $D$  and the goal probabilities  $\mathbf{p} = [p_0, \dots, p_D]$  when the probability of finding a goal before  $\delta_D$  is negligible.*

The assumption of DFS not finding a goal before  $\delta_D$  is not always realistic, but is for example satisfied in the grammar problems considered in Section 7 below.

## 6.1 DFS Analysis

The nodes  $\delta_0, \dots, \delta_D$  play a central role in the analysis of DFS runtime, since all the descendants of  $\delta_{n+1}$  will be explored before the descendants of  $\delta_n$  (excluding the  $\delta_{n+1}$  descendants). We say that *DFS explores from  $\delta_n$*  after DFS has explored all descendants of  $\delta_{n+1}$  and until all descendants of  $\delta_n$  have been explored. The general idea of the DFS analysis will be to count the number of nodes under each  $\delta_n$ , and to compute the probability that any of these nodes is a goal.

Some notation for this:

- Let the  $\delta_n$ -subgraph  $S_n = \{v : v \in \text{descendants}(\delta_n)\}$  be the set of nodes reachable from  $\delta_n$ , with cardinality  $|S_n| = \sum_{i=0}^D L(n, i)$ ,  $0 \leq n \leq D$ . Let  $S_{D+1} = \emptyset$  and let  $S_{-1}$  be a set of cardinality  $|S_{-1}| = |S_0| + 1 = \sum_{i=0}^D L(0, i) + 1$ .
- Let the  $\delta_n$ -explorables  $T_n = S_n \setminus S_{n+1}$  be the nodes explored from  $\delta_n$ .
- Let the number of level- $d$   $\delta_n$ -explorables  $A_{n,d} = L(n, d) - L(n+1, d)$  be the number of level  $d$  descendants of  $\delta_n$  that are not descendants of  $\delta_{n+1}$  for  $0 \leq n, d \leq D$ . The relation between  $T_n$  and  $A_{n,d}$  is the following:  $|T_n| = \sum_{i=n}^D A_{n,i}$ .

Let  $q_k = 1 - p_k$  for  $0 \leq k \leq D$ .

**Lemma 13.** *Consider a search problem with depth  $D$ , goal probabilities  $\mathbf{p}$ , and descendant counter  $L$ . The probability that the  $\delta_n$ -explorables  $T_n$  contains a goal is  $\tau_n := 1 - \prod_{k=0}^D q_k^{A_{n,k}}$ , and the probability that  $T_n$  contains the first goal is  $\phi_n := \tau_n \prod_{i=n+1}^D (1 - \tau_i)$ .*

*Proof.*  $\tau_n$  is 1 minus the probability of *not* hitting a goal at any level  $d$ ,  $n \leq d \leq D$ , since at each level  $d$ ,  $A_{n,d}$  probes are made when exploring from  $\delta_n$ .  $\square$

**Proposition 14** (Colliding branches expected DFS search time). *The expected DFS search time  $t_{\text{CB}}^{\text{DFS}}(D, \mathbf{p}, L)$  in a search problem with depth  $D$ , goal probabilities  $\mathbf{p}$ , and descendant counter  $L$  is bounded by*

$$t_{\text{CBL}}^{\text{DFS}}(D, \mathbf{p}, L) := \sum_{n=-1}^D |S_{n+1}| \phi_n \leq t_{\text{CB}}^{\text{DFS}}(D, \mathbf{p}, L) \leq \sum_{n=-1}^D |S_n| \phi_n := t_{\text{CBU}}^{\text{DFS}}(D, \mathbf{p}, l)$$

where  $\phi_{-1} = \bar{\Gamma} = 1 - \sum_{n=0}^D \phi_n$  is the probability that no goal exists.

The arithmetic mean  $\bar{t}_{\text{CB}}^{\text{DFS}}(D, \mathbf{p}, L) := (t_{\text{CBL}}^{\text{DFS}}(D, \mathbf{p}, L) + t_{\text{CBU}}^{\text{DFS}}(D, \mathbf{p}, L))/2$  between the bounds can be used for a single runtime estimate.

*Proof.* Let  $X$  be the DFS search time in a search problem with the features described above. The expectation of  $X$  may be decomposed as

$$\mathbb{E}[X] = P(\bar{\Gamma})\mathbb{E}[X \mid \bar{\Gamma}] + \sum_{n=0}^D P(\text{first goal in } T_n) \cdot \mathbb{E}[X \mid \text{first goal in } T_n]. \quad (1)$$

The conditional search time  $(X \mid \text{first goal in } T_n)$  is bounded by  $|S_{n+1}| \leq (X \mid \text{first goal in } T_n) \leq |S_n|$  for  $0 \leq n \leq D$ , since to find a goal DFS will search the entire  $\delta_{n+1}$ -subgraph  $S_{n+1}$  before finding it when searching the  $\delta_n$ -explorables  $T_n$ , but will not need to search more than the  $\delta_n$ -subgraph  $S_n = S_{n+1} \cup T_n$  (disregarding the few probes made ‘on the way down to’  $\delta_n$  (i.e. to  $T_n$ ); these probes were assumed negligible). The same bounds also hold with  $S_0$  and  $S_{-1}$  when no goal exists (recall that  $|S_{-1}| := |S_0| + 1$ ). Therefore the conditional expectation satisfies

$$|S_{n+1}| \leq \mathbb{E}[X \mid \text{first goal in } T_n] \leq |S_n| \quad (2)$$

for  $-1 \leq n \leq D$ . By Lemma 13, the probability that the first goal is among the  $\delta_n$ -explorables  $T_n$  is  $\phi_n$ , and the probability  $P(\bar{\Gamma})$  that no goal exists is  $\phi_{-1}$  by definition.

Substituting  $\phi_n$  and (2) into (1) gives the desired bounds for expected DFS search time  $\bar{t}_{\text{CB}}^{\text{DFS}}(D, \mathbf{p}, L) = \mathbb{E}[X]$ .  $\square$

The informativeness of the bounds of Proposition 14 depends on the dispersion of nodes between the different  $T_n$ ’s. If most nodes belong to one or a few sets  $T_n$ , the bounds may be almost completely uninformative. This happens in the special case of complete trees with branching factor  $b$ , where a fraction  $(b-1)/b$  of the nodes will be in  $T_0$ . The previous section derives techniques for these cases. The grammar problems investigated in Section 7 below show that the bounds may be relevant in more connected graphs, however.

## 6.2 BFS Analysis

The analysis of BFS only requires the descendant counter  $L(0, \cdot)$  with the first argument set to 0, and follows the same structure as Section 5.3.2. In contrast to the DFS bounds above, this analysis gives a precise expression for the expected runtime. The idea is to count the number of nodes in the upper  $k$  levels of the tree (derived from  $L(0, 0), \dots, L(0, k)$ ), and to compute the probability that they contain a goal. Let the *upper subgraph*  $U_k = \sum_{i=0}^{k-1} L(0, i)$  be the number of nodes above level  $k$ . When there is only a single goal level, Proposition 5 naturally generalises to the more general setting of this section.

**Lemma 15** (BFS runtime Single Goal Level). *For a search problem with depth  $D$  and descendant counter  $L$ , assume that the problem has a single goal level  $g$  with goal probability  $p_g$ , and that  $p_j = 0$  for  $j \neq g$ . When a goal exists and has position  $Y$  on the goal level, the BFS search time is:*

$$\begin{aligned} t_{\text{CB}}^{\text{BFS}}(g, p_g, L, Y) &= U_g + Y, \text{ with expected value} \\ t_{\text{CB}}^{\text{BFS}}(g, p_g, L \mid \Gamma_g) &= U_g + \text{tc}(p_g, L(0, g)) \end{aligned}$$

*Proof.* When a goal exists, BFS will explore all of the top of the tree until depth  $g - 1$  (that is,  $U_g$  nodes) and  $Y$  nodes on level  $g$  before finding the first goal. The expected value of  $Y$  is  $\text{tc}(p_g, L(0, g))$ .  $\square$

The probability that level  $k$  has a goal is  $P(\Gamma_k) = 1 - q_k^{L(0,k)}$ , and the probability that level  $k$  has the first goal is  $P(F_k) = P(\Gamma_k) \prod_{i=0}^{k-1} P(\bar{\Gamma}_i)$ . By the same argument as in Proposition 10, the following proposition holds.

**Proposition 16** (Branch Colliding Expected BFS Performance). *The expected number of nodes that BFS needs to search to find a goal in a search problem with depth  $D$ , goal probabilities  $\mathbf{p} = [p_0, \dots, p_D]$ ,  $\mathbf{p} \neq \mathbf{0}$ , and descendant counter  $L$  is*

$$t_{\text{CB}}^{\text{BFS}}(\mathbf{p}, L) = \sum_{k=0}^{D+1} P(F_k) t_{\text{CB}}^{\text{BFS}}(k, p_k, L \mid \Gamma_k)$$

where the goal probabilities have been extended with an extra element  $p_{D+1} = 1$ , and  $F_{D+1} = \bar{\Gamma}$  is the event that no goal exists.

For  $p_k = 0$ ,  $t_{\text{CB}}^{\text{BFS}}$  will be undefined, but this only occurs when  $P(F_k)$  is also 0. Proposition 14 and 16 give (rough) estimates of average BFS and DFS graph search time given the goal distribution  $\mathbf{p}$  and the structure parameter  $L$ . The results can be combined to make a decision whether to use BFS or DFS (Figure 5).

## 7 Grammar Problems

We now show how to apply the general theory of Section 6 to two concrete grammar problems. A *grammar problem* is a constructive search problem where nodes are strings over some finite alphabet  $B$ , and the neighbourhood relation is given by a set of production rules. *Production rules* are mappings  $x \rightarrow y$ ,  $x, y \in B^*$ , defining how strings may be transformed. For example, the production rule  $S \rightarrow Sa$  permits the string  $aSa$  to be transformed into  $aSaa$ . A grammar problem is defined by a set of production rules, together with a *starting string* and a *set of goal strings*. A *solution* is a sequence of production rule applications that transforms the starting string into a goal string. Many search problems can be formulated as grammar problems, with string representations of states modified by production rules. Their generality makes it *computationally undecidable* whether a given grammar problem has a solution or not. We here consider a simplified version where the search depth is artificially limited, and goals are distributed according to a goal probability vector  $p$ .

Grammar problems exhibit two features not present in the complete tree model. First, it is possible for branches of the grammar tree to ‘die’. This happens if no production rule is applicable to the string of the state. Second, often the same string can be produced by different sequences, which means that the grammar search graph in general is not a tree. The following subsections apply the theory of Section 6 of colliding branches to simple grammar problems.

### 7.1 Binary Grammar

Let  $\epsilon$  be the empty string. The *binary grammar* consists of two production rules,  $\epsilon \rightarrow a$  and  $\epsilon \rightarrow b$  over the alphabet  $B = \{a, b\}$ . The starting string is the empty

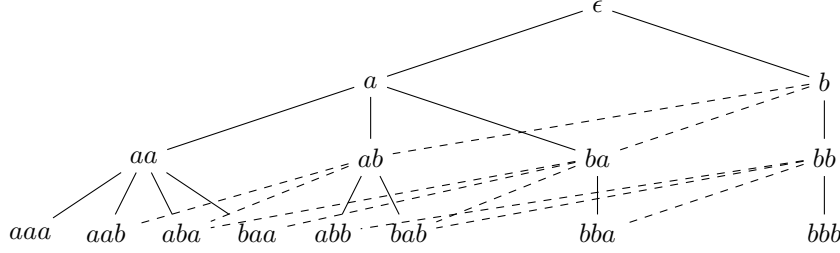


Figure 4: Graph of binary grammar problem with max depth  $D = 3$ . Contiguous lines indicate first discovery by DFS, and dashed lines indicate rediscoveries.

string  $\epsilon$ . A maximum depth  $D$  of the search graph is imposed, and strings on level  $k$  are goals with iid probability  $p_k$ ,  $0 \leq k \leq D$ . Since the left hand substring of both production rules is the empty string, both can always be applied at any place to a given string. The resulting graph is shown in Figure 4.

Consider a node  $v$  at level  $d$ . Its children are reached by either adding an  $a$  or by adding a  $b$ . Let  $\#a$  denote the number of  $a$ 's in  $v$ , and let  $\#b$  denote the number of  $b$ 's in  $v$ . Then  $\#a + 1$  distinct strings can be created by adding a  $b$ , and  $\#b + 1$  distinct strings can be created by adding an  $a$ . In total then,  $v$  will have  $(\#a + 1) + (\#b + 1) = d + 2$  children. Nodes further to the right will have more of their children previously discovered. The number of parents of a node is the number of contiguous  $a^i$  and  $b^j$  segments. For example,  $bbaaab$  have three segments  $bb$ - $aaa$ - $b$  and three parents  $baaab$ ,  $bbaab$  and  $bbaaa$ . A parent always differs from a child by the removal of one letter from one segment, and within a segment it is irrelevant which letter is removed.

The first node on level  $n$  that DFS reaches in the binary grammar problem is  $\delta_n = a^n$  for  $0 \leq n \leq D$ , assuming that the production rule  $\epsilon \rightarrow a$  is always used first by DFS. The following lemma derives an expression for the descendant counter  $L^{\text{BG}}$  required by Proposition 14. Incidentally, the number of level- $d$   $\delta_n$  explorables  $A_{n,d}$  (Section 6.1) gets an elegant form in the binary grammar problem.

**Lemma 17.** *For  $n < d$ , let  $L^{\text{BG}}(n, d) = |\{v : \text{level}(v) = d, v \in \text{descendants}(a^n)\}|$  be the number of nodes reachable from  $a^n$ , and let  $A_{n,d} = L^{\text{BG}}(n, d) - L^{\text{BG}}(n+1, d)$  be the number of descendants of  $a^n$  that are not descendants of  $a^{n+1}$ . Then  $L^{\text{BG}}(n, d) = \sum_{i=0}^{d-n} \binom{d}{i}$ , and  $A_{n,d} = \binom{d}{d-n}$ .*

*Proof.* The reachable nodes on level  $d$  that we wish to count are  $d-n$  levels below  $a^n$ . To reach this level we must add  $i \leq d-n$  number of  $b$ 's and  $d-n-i$  number of  $a$ 's to  $a^n$ . The number of length  $d$  strings containing exactly  $i$  number of  $b$ 's is  $\binom{d}{i}$  (we are choosing positions for the  $b$ 's non-uniquely with repetition among  $d-i+1$  possible positions). Summing over  $i$ , we obtain  $L^{\text{BG}}(n, d) = \sum_{i=0}^{d-n} \binom{d}{i}$ , and  $A_{n,d} = L^{\text{BG}}(n, d) - L^{\text{BG}}(n+1, d) = \binom{d}{d-n}$ .  $\square$

**Corollary 18** (Expected Binary Grammar BFS Search Time). *The expected BFS search time  $t_{\text{BG}}^{\text{DFS}}(\mathbf{p})$  in a Binary Grammar Problem of depth  $D$  with goal probabilities  $\mathbf{p} = [p_0, \dots, p_D]$  is*

$$t_{\text{BG}}^{\text{BFS}}(\mathbf{p}) = t_{\text{CB}}^{\text{BFS}}(\mathbf{p}, L^{\text{BG}}).$$

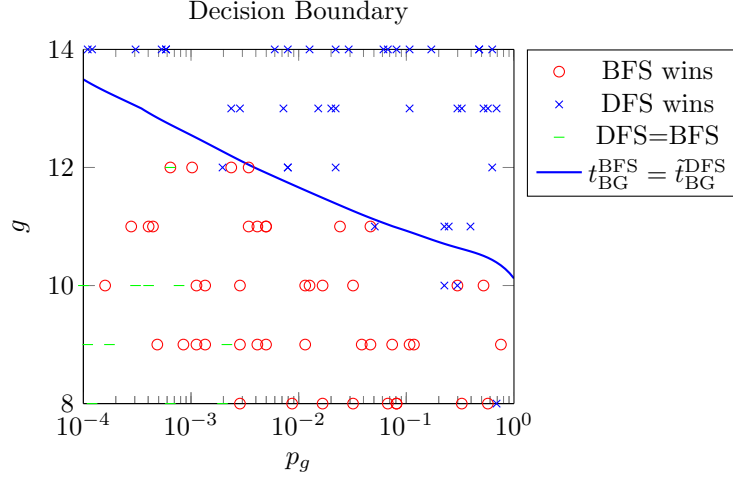


Figure 5: The decision boundary predicted by Corollary 18 and 19, together with empirical outcomes of BFS and DFS search time. The scattered points are based on 100 independently generated binary grammar problems of depth  $D = 14$  with uniformly sampled (single) goal level  $g \in [8, 14] \cap \mathbb{N}$  and  $\log(p_g) \in [-4, 0]$ . DFS benefits from a deeper goal level and higher goal probability compared to BFS. The decision boundary gets 87% of the instances correct.

**Corollary 19** (Expected Binary Grammar DFS Search Time). *The expected DFS search time  $\tilde{t}_{\text{BG}}^{\text{DFS}}(D, \mathbf{p})$  in a binary grammar problem of depth  $D$  with goal probabilities  $\mathbf{p} = [p_0, \dots, p_D]$  is bounded between  $t_{\text{BGL}}^{\text{DFS}}(D, \mathbf{p}) := t_{\text{CBL}}^{\text{DFS}}(D, \mathbf{p}, L^{\text{BG}})$  and  $t_{\text{BGU}}^{\text{DFS}}(D, \mathbf{p}) := t_{\text{CBU}}^{\text{DFS}}(D, \mathbf{p}, L^{\text{BG}})$ , and is approximately*

$$\tilde{t}_{\text{BG}}^{\text{DFS}}(D, \mathbf{p}) := \tilde{t}_{\text{CB}}^{\text{DFS}}(D, \mathbf{p}, L^{\text{BG}}).$$

*Proof of Corollary 18 and 19.* Direct application of Lemma 17, and Proposition 16 and 14 respectively.  $\square$

The bounds are plotted for a single goal level in Figure 5 and 6.

## 7.2 Random Grammar

The Random Grammar Problem has alphabet  $B = \{S, a, b\}$  and start string  $S$ . The *production rules* always include  $S \rightarrow \epsilon$  (with  $\epsilon$  denoting the empty string) plus a random subset of the *adding rules*  $S \rightarrow Sa$ ,  $S \rightarrow Sb$ ,  $S \rightarrow aS$ ,  $S \rightarrow bS$ , and a random subset of the *moving rules*  $Sa \rightarrow aS$ ,  $Sb \rightarrow bS$ ,  $aS \rightarrow Sa$ , and  $bS \rightarrow Sb$ . Only strings containing no  $S$  can be goal nodes. As usual, a maximum depth  $D$  and a goal probability vector  $p = [p_0, \dots, p_D]$  are given.

For simplified analysis, we will abuse notation the following way. We will consider  $S$ -less nodes to be one level higher than they actually are. For example, we will consider  $a$  to be on level 1, although it is technically on level 2 or lower (e.g. reached by the path  $S \rightarrow Sa$ ,  $S \rightarrow \epsilon$ ). A slight modification of BFS and DFS makes them always check the  $S$ -less child first (which is always child-less in turn), which means the change will only slightly affect search time. We will still consider  $\delta_n = Sa^n$  whenever  $S \rightarrow Sa$  is among the production rules, however.

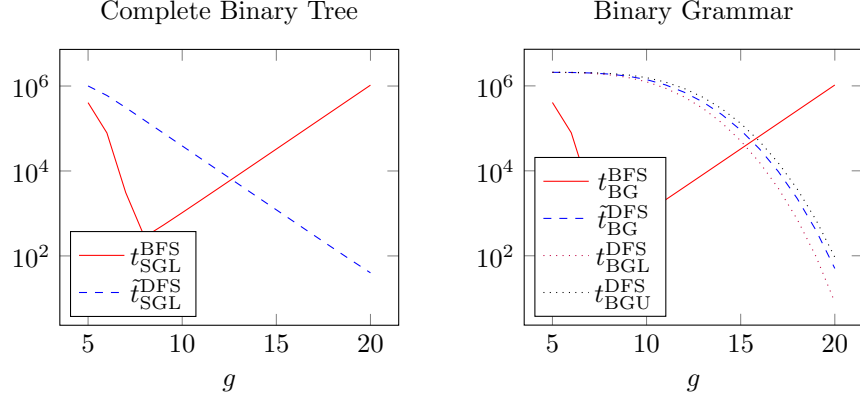


Figure 6: The expected search time of BFS and DFS as a function of a single goal level  $g$  with goal probability  $p_g = 0.05$  in a tree of depth  $D = 20$ . BFS has the advantage when the goal is in the higher regions of the graph, although at first the probability that no goal exists heavily influences both BFS and DFS search time. The greater connectivity of the graph in the binary grammar problem permits DFS to spend more time in the lower regions before backtracking, compared to the complete binary tree analysed in the previous section. This penalises DFS runtime when the goal is not in the very lowest regions of the tree. BFS behaviour is identical in both models.

The general case of when a random set of production rules are used is explored experimentally in Section 9. The special case of a binary tree arises when none of the moving rules are used, and either only the first two or only the last two of the addition rules. The analysis of Section 5.2 applies to this case. The special case when all rules are present can be analysed analytically by the means of Section 6. We will call this case the full grammar problem.

### 7.3 Full Grammar

The search graph of the full grammar problem is shown in Figure 7 (edges induced by moving rules are not shown). Since there are four adding rules that can be applied to each node, each node will have four children. Typically, when we move further to the right in the tree, more children will already have been discovered.

The full grammar problem can be analysed by a reduction to a binary grammar problem with the same parameters  $D$  and  $\mathbf{p}$ . Assign to each string  $v$  of the binary grammar problem the set of strings that only differ from  $v$  by (at most) an extra  $S$ . We call such sets *node clusters*. For example,  $\{a, Sa, aS\}$  constitutes the node cluster corresponding to  $a$ . Due to the abusing of levels for the  $S$ -less strings, all members of a cluster appear on the same level in the full grammar problem (the level is equal to the number of  $a$ 's and  $b$ 's). The level is also the same as the corresponding string in the binary grammar problem.

**Lemma 20** (Binary Grammar Reduction). *For every  $n, d, n \leq d$ , the descendant counter  $L^{\text{FG}}$  of the full grammar problem is  $L^{\text{FG}}(n, d) = (d + 2)L^{\text{BG}}(n, d)$ .*

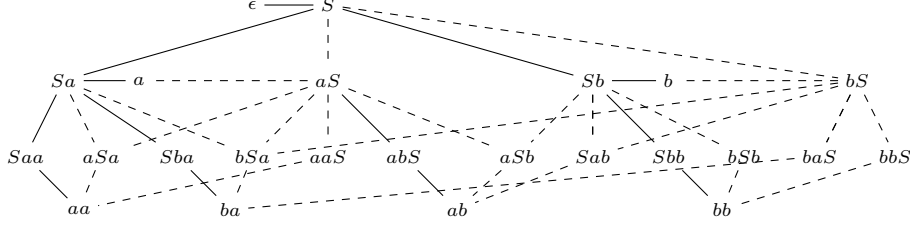


Figure 7: Search graph for the Grammar problem until level 2. Connections induced by moving rules are not displayed. Contiguous lines indicate the first discovery of a child by DFS and dashed lines indicate rediscoveries.

*Proof.*  $L^{\text{BG}}(n, d)$  counts the level  $d$  descendants of  $a^n$  in the binary grammar problem (BGP), and  $L^{\text{FG}}(n, d)$  counts the level  $d$  descendants of  $Sa^n$  in the full grammar problem (FGP). The node  $u$  is a child of  $v$  in BGP iff the members of the  $u$  node cluster are descendants of  $Su$ . Therefore the node clusters on level  $d$  descending from  $Sa^n$  in FGP correspond to the BGP nodes descending from  $a^n$ . At level  $d$ , each node cluster contains  $d + 2$  nodes.  $\square$

**Corollary 21** (Expected Full Grammar BFS Search Time). *The expected BFS search time  $\tilde{t}_{\text{FG}}^{\text{DFS}}(\mathbf{p})$  in a full grammar problem of depth  $D$  with goal probabilities  $\mathbf{p} = [p_0, \dots, p_D]$  is*

$$t_{\text{FG}}^{\text{BFS}}(\mathbf{p}) := t_{\text{CB}}^{\text{BFS}}(\mathbf{p}, L^{\text{FG}}).$$

**Corollary 22** (Expected Full Grammar DFS Search Time). *The expected DFS search time  $\tilde{t}_{\text{FG}}^{\text{DFS}}(D, \mathbf{p})$  in a full grammar problem of depth  $D$  with goal probabilities  $\mathbf{p} = [p_0, \dots, p_D]$  is bounded between  $t_{\text{FGL}}^{\text{DFS}}(D, \mathbf{p}) := t_{\text{CBL}}^{\text{DFS}}(D, \mathbf{p}, L^{\text{FG}})$  and  $t_{\text{FGU}}^{\text{DFS}}(D, \mathbf{p}) := t_{\text{CBU}}^{\text{DFS}}(D, \mathbf{p}, L^{\text{FG}})$ , and is approximately*

$$\tilde{t}_{\text{FG}}^{\text{DFS}}(D, \mathbf{p}) := \tilde{t}_{\text{CB}}^{\text{DFS}}(D, \mathbf{p}, L^{\text{FG}}).$$

*Proof of Corollary 21 and 22.* Direct application of Lemma 20, and Proposition 16 and 14 respectively.  $\square$

## 8 Experimental verification

To verify the analytical results, we have implemented the models Sections 5–7 in Python 3 using the `graph-tool` package (Peixoto, 2015). The data reported in Tables 1–3 is based on an average of 1000 independently generated search problems with depth  $D = 14$ .

- The first number in each box is the empirical average,
- the second number is the analytical estimate, and
- the third number is the percentage error of the analytical estimate.

For certain parameter settings, there is only a small chance ( $< 10^{-3}$ ) that there are no goals. In such circumstances, all 1000 generated search graphs typically inhabit a goal, and so the empirical search times will be comparatively small. However, since a tree of depth 14 has about  $2^{15} \approx 3 \cdot 10^5$  nodes (and



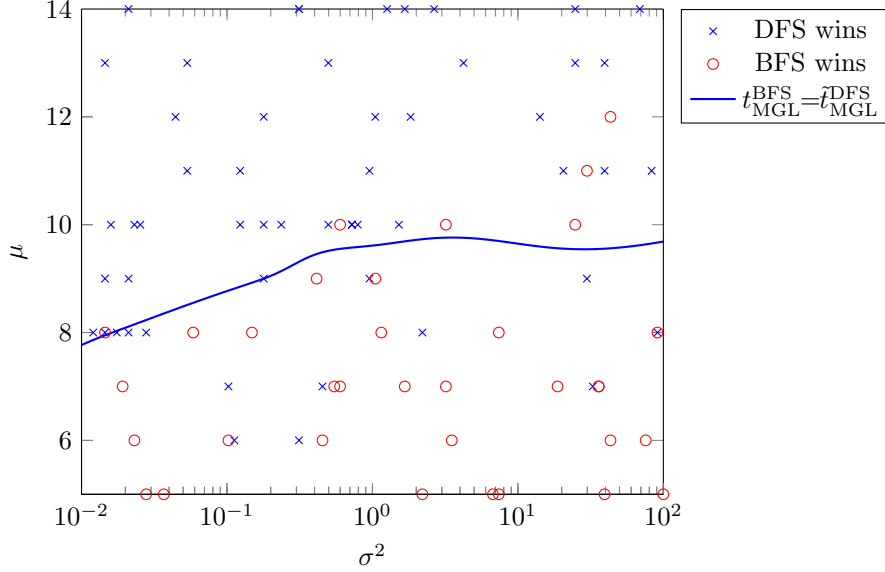


Figure 8: The decision boundary for the Gaussian tree given by Proposition 9 and 10, together with empirical outcomes of BFS vs. DFS winner. The scattered points are based on 100 independently generated problems with depth  $D = 14$  and uniformly sampled parameters  $\mu \in [5, 14] \cap \mathbb{N}$  and  $\log(\sigma^2) \in [-2, 2]$ . The most deciding feature is the goal peak  $\mu$ , but DFS also benefits from a smaller  $\sigma^2$ . The decision boundary gets 74% of the winners correct.

a search algorithm must search through all of them in case there is no goal), the rarely occurring event of no goal can still influence the *expected* search time substantially. To avoid this sampling problem, we have ubiquitously discarded all instances where no goal is present, and compared the resulting averages to the analytical expectations *conditioned on at least one goal being present*.

To develop a concrete instance of the multi goal level model we consider the special case of *Gaussian goal probability vectors*, with two parameters  $\mu$  and  $\sigma^2$ . For a given depth  $D$ , the goal probabilities are given by

$$p_i = \min \left\{ \frac{1}{20\sqrt{\sigma^2}} e^{(i-\mu)^2/\sigma^2}, \frac{1}{2} \right\}.$$

The parameter  $\mu \in [0, D] \cap \mathbb{N}$  is the *goal peak*, and the parameter  $\sigma^2 \in \mathbb{R}^+$  is the *goal spread*. The factor  $1/20$  is arbitrary, and chosen to give an interesting dynamics between searching depth-first and breadth-first. No  $p_i$  should be greater than  $1/2$ , in order to (roughly) satisfy the assumption of Proposition 10. We call this model the *Gaussian binary tree*.

**Complete Tree** The accuracy of the predictions of Proposition 5 and 6 are shown in Table 1, and the accuracy of Proposition 9 and 10 in Table 2. The relative error is always small for BFS ( $< 10\%$ ). For DFS the error is generally within 20%, except when the search time is small ( $< 35$  probes), in which case the absolute error is always small. The decision boundary of Proposition 7 is

$g \backslash p_g$	0.001	0.01	0.1
5		46.33	40.01
		46.64	39.86
		0.7 %	0.4 %
8	369.5	332.8	264.6
		378.0	265.0
		2.3 %	0.2 %
11	2748	2143	2057
		2744	2057
		0.1 %	0. %
14	17 360	16 480	16 390
		17 380	16 390
		0.1 %	0. %

(a) BFS single goal level

$g \backslash p_g$	0.001	0.01	0.1
5		14 680	8206
		15 000	8053
		2.2 %	1.9 %
8	14 530	9833	1105
		15 620	1154
		7.5 %	4.5 %
11	11 200	1535	152.3
		11 140	146.0
		0.5 %	4.1 %
14	1971	208.8	30.57
		2000	20.00
		1.4 %	35 %

(b) DFS single goal level

Table 1: BFS and DFS performance in the single goal level model with depth  $D = 14$ , where  $g$  is the goal level and  $p_g$  the goal probability. Each box contains empirical average/analytical expectation/error percentage.

$\mu \backslash \sigma$	0.1	1	10	100
5	37.24	43.75	90.87	225.1
	37.04	41.55	83.72	210.8
	0.5 %	5.0 %	7.9 %	6.4 %
8	261.2	171.9	119.6	212.0
	261.3	173.4	119.8	211.0
	0. %	0.9 %	0.2 %	0.5 %
11	2049	953.0	303.9	249.5
	2050	953.0	305.0	247.5
	0. %	0. %	0.3 %	0.8 %
14	16 210	5159	968.5	332.9
	16 150	5136	960.6	329.7
	0.4 %	0.4 %	0.8 %	0.9 %

(a) BFS multi goal level

$\mu \backslash \sigma$	0.1	1	10	100
5	5374	8572	3405	385.8
	5949	10 070	3477	379.1
	11 %	18 %	2.1 %	1.7 %
8	677.3	1234	454.6	252.6
	743.6	1259	473.6	260.0
	9.8 %	2.1 %	4.2 %	2.9 %
11	97.38	168.1	117.4	210.0
	92.95	157.4	106.7	211.7
	4.5 %	6.4 %	9.1 %	0.8 %
14	24.00	43.38	81.75	213.6
	11.62	32.89	74.46	205.0
	52 %	24 %	8.9 %	4.0 %

(b) DFS multi goal level

Table 2: BFS and DFS performance in Gaussian binary trees with depth  $D = 14$ . Each box contains empirical average/analytical expectation/error percentage.

shown in Figure 2, and the decision boundary of Proposition 9 vs. 10 is shown in Figure 8. These boundary plots show that the analysis generally predict the correct BFS vs. DFS winner.

**Grammar** The binary grammar model of Section 7.1 serves to verify the general estimates of Proposition 14 and 16. The results are shown in Table 3. The estimates for BFS are accurate as usual ( $< 3\%$  error). With few exceptions, the lower and the upper bounds  $t_{\text{BGL}}^{\text{DFS}}$  and  $t_{\text{BGU}}^{\text{DFS}}$  of Corollary 19 for DFS differ by at most 50% on the respective sides from the true (empirical) average. The arithmetic mean  $\bar{t}_{\text{BG}}^{\text{DFS}}$  often give surprisingly accurate predictions ( $< 4\%$ ) except when  $t_{\text{BGL}}^{\text{DFS}}$  and  $t_{\text{BGU}}^{\text{DFS}}$  leave wide margins as to the expected search time (when  $g = 14$ , the margin is up to 84% downwards and 125% upwards). Even then, the  $\bar{t}_{\text{BG}}^{\text{DFS}}$  error remains within 30%.

## 9 Empirical Predictions

The Random Grammar model of Section 7.2 exhibits a rich variety of topological features such as properties of the branching factor distribution. It is an interesting

$g \backslash p_g$	0.001	0.01	0.1
5		46.74	40.53
		46.64	39.86
		0.2 %	1.7 %
8	375.7 378.0 0.6 %	332.5	265.7
		333.9	265.0
		0.4 %	0.3 %
11	2751 2744 0.3 %	2145	2058
		2147	2057
		0.1 %	0. %
14	17 370 17 380 0.1 %	16 480	16 390
		16 480	16 390
		0. %	0. %

(a) BFS  $t_{\text{BG}}^{\text{BFS}}$ 

$g \backslash p_g$	0.001	0.01	0.1
5		30 910	27 840
		31 370	30 190
		1.5 %	8.4 %
8	28 000 27 410 2.1 %	25 160	15 490
		24 420	15 200
		2.9 %	1.9 %
11	17 280 16 790 2.9 %	5932	1815
		5806	1788
		2.1 %	1.5 %
14	1304 1522 17 %	122.1	25.60
		164.6	20.06
		35 %	22 %

(b) Average DFS  $\bar{t}_{\text{BG}}^{\text{DFS}}$ 

$g \backslash p_g$	0.001	0.01	0.1
5		30 910	27 840
		30 710	29 080
		0.7 %	4.5 %
8	28 000 25 740 8.1 %	25 160	15 490
		22 150	12 070
		12 %	22 %
11	17 280 14 160 18 %	5932	1815
		3822	918.6
		36 %	49 %
14	1304 808.8 38 %	122.1	25.60
		54.12	3.990
		56 %	84 %

(c) Lower DFS  $t_{\text{BGL}}^{\text{DFS}}$ 

$g \backslash p_g$	0.001	0.01	0.1
5		30 910	27 840
		32 020	31 290
		3.6 %	12 %
8	28 000 29 080 3.8 %	25 160	15 490
		26 690	18 340
		6.1 %	18 %
11	17 280 19 410 12 %	5932	1815
		7790	2657
		31 %	46 %
14	1304 2236 72 %	122.1	25.60
		275.1	36.12
		125 %	41 %

(d) Upper DFS  $t_{\text{BGU}}^{\text{DFS}}$ 

Table 3: BFS and DFS performance in binary grammars of depth  $D = 14$ . Empirical DFS performance is compared to the upper and lower bounds of Corollary 19, as well as their arithmetic average. In these experiments, goals are distributed on a single goal level  $g$  with goal probability  $p_g$ . The BFS estimates  $t_{\text{BG}}^{\text{BFS}}$  are highly accurate, and the averaged DFS estimates  $\bar{t}_{\text{BG}}^{\text{DFS}}$  are mostly accurate. Each box contains empirical average/analytical expectation/error percentage.

question to what extent such features can be used to predict whether BFS or DFS is the better search method. However, it is harder to approach analytically, due to its many cases not the least. We therefore approached this problem empirically.

We generated a data set with 1827 randomly sampled random grammars. The sampling was done uniformly from the following sets: First sample a number of rules  $r \in [4, 8] \cap \mathbb{N}$ , then a random size  $r$  subset of the 8 possible production rules. Also sample maximum depth  $D \in [11, 15] \cap \mathbb{N}$ . Sample a number of goals  $n \in [3, 5D] \cap \mathbb{N}$ . Sample  $n$  times a level  $k \in [1, D] \cap \mathbb{N}$  and a node on level  $k$ ; make the sampled node a goal.

We trained a Support Vector Machine from the `scikit-learn` package (Pedregosa et al., 2011) with a (Gaussian) Radial Basis Kernel to predict whether BFS or DFS would find a first goal faster. The features we predicted from were

- Mean branching factor
- Standard deviation of branching factor
- Number of rules
- Maximum depth

The best parameter settings for the support vector machine were  $C = 1000$ ,  $degree = 3$ ,  $\gamma = 0.1$ . Against a cross-validation data set, the trained support vector classifier got the BFS/DFS winner correct in 63% of the cases, in a dataset where DFS won 55% of the time. The results give an early indication that it may be possible to predict the best search method based solely on locally estimable features of the search graph.

## 10 Discussion

Search and optimisation problems appears in different flavors throughout the field of artificial intelligence; in planning, problem solving, games, and learning. Therefore even minor improvements to search performance can potentially lead to gains in many aspects of intelligent systems. It is even possible to equate intelligence with (Bayesian expectimax) optimisation performance (Legg and Hutter, 2007).

**Summary.** In this report we have derived analytical results for expected runtime performance. Section 5 focused on BFS and DFS *tree search* where explored nodes were not remembered. A vector  $\mathbf{p} = (p_1, \dots, p_D)$  described *a priori* goal probabilities for the different levels of the tree. This concrete but general model of goal distribution allowed us to calculate approximate closed-form expression of both BFS and DFS average runtime. Earlier studies have only addressed *worst-case* runtimes: Knuth (1975) and followers for DFS; Korf et al. (2001) and followers for IDA\*, effectively a generalised version of BFS.

Section 6 and 7 generalised the model of Section 5 to non-tree graphs. In addition to the goal probability vector  $\mathbf{p}$ , the graph search analysis required additional structural information in the form of a descendant counter  $L$ . The graph search estimates also took the form of less precise bounds. The analysis of Section 6 does not supersede the analysis in Section 5, as the bounds of

Section 6 become uninformative when the graph is a tree. The results are generally consistent with empirical reality.

**Conclusions and Outlook.** The value of the results are at least twofold. They offer a concrete means of deciding between BFS and DFS given some rough idea of the location of the goal (and the graph structure). To make the results more generally usable, automatic inference of model parameters would be necessary; primarily of goal distribution  $p$  and graph structure  $L$ . (The depth  $D$  will often be set by the searcher itself, and perhaps be iteratively increased.) There is good hope that the descendant counter  $L$  can be estimated online from the local sample obtained during search, similar to (Knuth, 1975). The goal distribution is likely to prove more challenging, but resembles the automatic creation of heuristic functions, so techniques such as *relaxed problems* could well prove useful (Pearl, 1984). Estimates of goal distribution could possibly also be inferred from a heuristic function.

The results also offer theoretical insight into BFS and DFS performance. As BFS and DFS are in a sense the most fundamental search operations, we have high hopes that our results and techniques will prove useful as building blocks for analysis of more advanced search algorithms as well. For example, A\* and IDA\* may be viewed as generalisations of BFS, and Beam Search and Greedy Best-First as generalisations of DFS.

## References

- Arbelaez Rodriguez, A. (2011). *Learning During Search*. Phd thesis, University of Paris-Sud.
- Blum, C. and Roli, A. (2003). Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. *ACM Computing Surveys*, 35(3):268–308.
- Burke, E., Hart, E., Kendall, G., Newall, J., Ross, P., and Schulenburg, S. (2003). Hyper Heuristics: an emerging direction in modern search technology. In Glover, F. and Kochenberger, G. A., editors, *Handbook of Metaheuristics*, pages 457–474. Springer.
- Burke, E. K., Gendreau, M., Hyde, M., Kendall, G., Ochoa 1ã, G., Zcan, E., and Qu, R. (2013). Hyper-heuristics: a survey of the state of the art. *Journal of the Operational Research Society*, 64:1695–1724.
- Chen, P. C. (1992). Heuristic Sampling: A Method for Predicting the Performance of Tree Searching Programs. *SIAM Journal on Computing*, 21(2):295–315.
- Dejong, G. and Mooney, R. (1986). Explanation-based Learning: An Alternative View. *Machine Learning*, 1(2):145–176.
- Diestel, R. (2006). *Graph Theory (Graduate Texts in Mathematics)*.
- Edelkamp, S. and Schrödl, S. (2012). *Heuristic Search*. Morgan Kaufmann Publishers Inc.

- Etzioni, O. (1993). Acquiring search-control knowledge via static analysis. *Artificial Intelligence*, 62:255–301.
- Everitt, T. and Hutter, M. (2015a). Analytical Results on the BFS vs. DFS Algorithm Selection Problem. Part I: Tree Search. In *28th Australian Joint Conference on Artificial Intelligence*.
- Everitt, T. and Hutter, M. (2015b). Analytical Results on the BFS vs. DFS Algorithm Selection Problem. Part II: Graph Search. In *28th Australian Joint Conference on Artificial Intelligence*.
- Fink, E. (1998). How to Solve It Automatically: Selection Among Problem-Solving Methods. In *Proceedings of the Fourth International Conference on Artificial Intelligence Planning Systems*, pages 128–136.
- Gomes, C. P., Gomes, C. P., Selman, B., Selman, B., Crato, N., and Crato, N. (1997). Heavy-tailed distributions in combinatorial search. *Principles and Practice of Constraint Programming-CP97*, 1330:121–135.
- Haim, S. and Walsh, T. (2008). Online estimation of SAT solving runtime. In *Theory and Applications of Satisfiability Testing*, volume 4996 LNCS, pages 133–138. Springer.
- Hoos, H. H. (2012). Programming by optimization. *Communications of the ACM*, 55(2):70.
- Hutter, F., Xu, L., Hoos, H. H., and Leyton-Brown, K. (2014). Algorithm runtime prediction: Methods & evaluation. *Artificial Intelligence*, 206(1):79–111.
- Kilby, P., Slaney, J., Thiébaux, S., and Walsh, T. (2006). Estimating Search Tree Size. In *Proc. of the 21st National Conf. of Artificial Intelligence, AAAI, Menlo Park*.
- Knuth, D. E. (1975). Estimating the efficiency of backtrack programs. *Mathematics of Computation*, 29(129):122–122.
- Korf, R. E., Reid, M., and Edelkamp, S. (2001). Time complexity of iterative-deepening-A\*. *Artificial Intelligence*, 129(1-2):199–218.
- Kotthoff, L. (2014). Algorithm Selection for Combinatorial Search Problems: A Survey. *AI Magazine*, pages 1–17.
- Kullmann, O. (2008). Fundamentals of Branching Heuristics: Theory and Examples. Technical report, Computer Science Department, Swansea University, UK.
- Leckie, C. and Zukerman, I. (1998). Inductive learning of search control rules for planning. *Artificial Intelligence*, 101:63–98.
- Legg, S. and Hutter, M. (2007). Universal Intelligence. *Minds & Machines*, 17(4):391–444.
- Lelis, L. H. S., Otten, L., and Dechter, R. (2013). Predicting the size of Depth-first Branch and Bound search trees. *IJCAI International Joint Conference on Artificial Intelligence*, pages 594–600.

- Minton, S. (1988). *Learning Search Control Knowledge: An Explanation-Based Approach*. Kluwer Academic Publishers.
- Minton, S. (1990). Quantitative results concerning the utility of explanation-based learning. *Artificial Intelligence*, 42(1990):363–391.
- Mitchell, T., Keller, R., and Kedar-Cabelli, S. (1986). Explanation-based generalization: A unifying view. *Machine Learning*, 1(1):47–80.
- Pearl, J. (1984). *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Addison-Wesley.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825—2830.
- Peixoto, T. P. (2015). The graph-tool python library. *figshare*.
- Purdum, P. W. (1978). Tree Size by Partial Backtracking. *SIAM Journal on Computing*, 7(4):481–491.
- Rice, J. R. (1975). The algorithm selection problem. *Advances in Computers*, 15:65–117.
- Rokicki, T. and Kociemba, H. (2013). The diameter of the rubiks cube group is twenty. *SIAM Journal on Discrete Mathematics*, 27(2):1082–1105.
- Ross, P., Schulenburg, S., Marin-Blazquez, J. G., and Hart, E. (2002). Hyperheuristics: learning to combine simple heuristics in bin-packing problems. In *GECCO’02*.
- Russell, S. J. and Norvig, P. (2010). *Artificial intelligence: a modern approach*. Prentice Hall, third edition.
- Smith-Miles, K. and Lopes, L. (2012). Measuring instance difficulty for combinatorial optimization problems. *Computers and Operations Research*, 39(5):875–889.
- Thompson, C. (2011). *Metareasoning about propagators for constraint satisfaction*. Phd thesis, University of Saskatchewan.
- Wu, Z. and Preciado, V. M. (2013). Laplacian Spectral Properties of Graphs from Random Local Samples.
- Zahavi, U., Felner, A., Burch, N., and Holte, R. C. (2010). Predicting the performance of IDA\* using conditional distributions. *Journal of Artificial Intelligence Research*, 37:41–83.



## A List of notation

$P$	Probability
$X, Y$	Random variables
$\mathbb{E}[\cdot]$	Expectation of a random variable
$O$	Big-O notation
$EC$	Edge cost
$h$	Heuristic function
$g$	Accumulated path cost from start node
$Q$	Objective function
$D$	Maximum depth of search space
$p_g$	Goal probability at a single goal level $g$
$p_k$	Goal probability for a level $k$
$\mathbf{p}$	Vector of probabilities for multiple goal levels
$\mu, \sigma^2$	Goal peak and goal spread in Gaussian binary tree
$\Gamma$	Probability that a goal exists
$\Gamma_k$	Probability that level $k$ has a goal
$F_k$	Probability that level $k$ has the first goal
$t_{\text{SGL}}^{\text{BFS}}, \tilde{t}_{\text{SGL}}^{\text{DFS}}$	Expected BFS search time and approximate expected DFS search time in a complete tree with a single goal level
$t_{\text{MGL}}^{\text{BFS}}, \tilde{t}_{\text{MGL}}^{\text{DFS}}$	Expected BFS search time and approximate expected DFS search time in a complete tree with multiple goal levels
$t_{\text{CB}}^{\text{BFS}}, \tilde{t}_{\text{CB}}^{\text{DFS}}$	Expected BFS search time and approximate expected DFS search time in a graph with colliding branches
$t_{\text{BG}}^{\text{BFS}}, \tilde{t}_{\text{BG}}^{\text{DFS}}$	Expected BFS search time and approximate expected DFS search time in the binary grammar problem
$t_{\text{FG}}^{\text{BFS}}, \tilde{t}_{\text{FG}}^{\text{DFS}}$	Expected BFS search time and approximate expected DFS search time in the full grammar problem
$\delta_n$	The first node on level $n$ reached by DFS
$L(n, d)$	Descendant counter, counting the number of level $d$ descendants reachable from $\delta_n$
$L^{\text{FG}}, L^{\text{BG}}$	Descendant counters for the binary grammar problem and the full grammar problem
$A_{n,d}$	Number of nodes reachable from $\delta_n$ not reachable from $\delta_{n+1}$
$S_n$	Descendants of $\delta_n$
$T_n$	Descendants of $\delta_n$ that are not descendants of $\delta_{n+1}$
$U_n$	The number of nodes above level $n$ .
$\tau_n$	The probability that $T_n$ contains a goal (Lemma 13)
$\phi_n$	The probability that $T_n$ inhabits the first goal
$b$	Branching factor
$\epsilon$	Empty string

## B List of search problems

The following is an incomplete list of problems naturally modelled as search problems, organised by type.

- Puzzles
  - N-puzzle
  - Instant insanity (Knuth 1975)
  - Eternity II (Assembly puzzles)
- Infinite
  - Grammar (aka Production system; simpler PSVN)
  - STRIPS planning (PDDL language)
  - Root-factorial (Knuth)
- Real-world problems
  - MDP
  - SAT
  - VLSI chip design (cell layout, channel routing)
  - Robot navigation (continuous)
  - Route finding
  - Tour finding
  - Assembly sequencing
  - Protein design
  - Jobshop (who does what task when)
- Other
  - Towers of Hanoi
  - Cannibal-missionary
  - Sokoban
  - Rubik's cube
  - Sudoku
  - Knight jumping
  - N-queens
  - Belief state (in a deterministic, partially observable world)
  - Quasigroup completion problem (Gomes et al., 1997), naturally CSP
  - Counterfeit coin problem (Pearl, 1988)

## C List of Topological features

The neighbourhood relation  $N$  induces a topology on the state space  $S$ . The goal of this study is to closer investigate how topological features of the search graph affects search performance. Following is a list of potentially useful features, where +, ?, - ranks the features by *a priori* likelihood of being useful. Please refer to (Diestel, 2006) for a standard reference on graph theory and for explanations and definitions of below terms.

- Problem type
  - + Directed/undirected graph
- State space
  - + number of nodes (finite or infinite)
  - number of edges
- Graph structure
  - bipartite/ $k$ -partite
  - + clique size distribution
  - + clique covering number (how many cliques are required to cover the graph?)
  - chordal (every cycle of length  $\geq 4$  has a "chord")
  - stability number (greatest number of non-connected nodes)
  - + degree (max/min/average)
  - min cycle (girth)
  - max cycle (circumference)
  - diameter =  $\max_{x,y} \text{distance}(x,y)$
  - radius =  $\min_x \max_y \text{distance}(x,y)$  ( $x$  is 'the centre' of the graph).  
The radius satisfies  $\text{radius}(G) \leq \text{diameter}(G) \leq 2 \cdot \text{radius}(G)$ .
- Tree structure
  - + height
  - + max width
  - + width as a function of depth
- Branching factor
  - + Distribution (possibly as a function of depth)
  - + Max
  - + Effective (when using heuristic)
- Path length
  - + Max/average path length from origin, repeating allowed/disallowed
- Path redundancy
  - +  $k$ -connectedness (every node pair have at least  $k$  independent paths)
  - $l$ -edge-connectedness (no cutset of  $l$  edges)
  - + distribution of connectedness/revisiting frequency

- Matrix properties
  - + Spectrum (Spectral Graph Theory)
- Edge space
  - ? Cyclomatic number (defined in (Diestel, 2006, p. 24) as the dimension of the space formed by cycles – a subspace of the space of edges under "symmetric difference")
  - Cut space dimension

Many properties of the search graph can be estimated from local samples. For example, Wu and Preciado (2013) show that the spectrum of the graph (i.e., the eigenvalues of the matrix representation of the graph) can be estimated this way.